

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

ISOLATION AND IDENTIFICATION OF NOVEL POLYMERASES

Field of the Invention

This invention relates to newly identified polynucleotides, polypeptides encoded by such polynucleotides, the use of such polynucleotides and polypeptides, as well as the production and isolation of such polynucleotides and polypeptides. More particularly, the polypeptides of the present invention have been identified as polymerases

Background of the Invention

Thermophilic bacteria have received considerable attention as sources of highly active and thermostable enzymes. Recently, the most extremely thermophilic organotrophic eubacteria presently known have been isolated and characterized. These bacteria, which belong to the genus *thermotoga*, are fermentative microorganisms metabolizing a variety of carbohydrates (Huber, R. and Stetter, K.O., in Ballows, et al., (Ed.), The Prokaryotes, 2nd Ed., Springer-Verlag, New York, pgs. 3809-3819 (1992)).

In Huber *et al.*, 1986, Arch. Microbiol. 144:324-333, the isolation of the bacterium *Thermotoga maritima* is described. *T. maritima* is a eubacterium that is strictly anaerobic, rod-shaped, fermentative, hyperthermophilic, and grows between 55°C and 90°C, with an optimum growth temperature of about 80°C. This eubacterium has been isolated from geothermally heated sea floors in Italy and the Azores. *T. maritima* cells have a sheath-like structure and monotrichous flagellation. *T. maritima* is classified in the eubacterium kingdom by virtue of having murein and fatty acid-containing lipids, diphtheria-toxin-resistant elongation factor 2, an RNA polymerase subunit pattern, and sensitivity to antibiotics.

Since, to date, most organisms identified from the archaeal domain are thermophiles or hyperthermophiles, archaea are also considered a fertile source of thermophilic enzymes.

Summary of the Invention

The present invention provides polynucleotides and polypeptides encoded thereby which have been identified as polymerase enzymes. In accordance with one aspect of the present invention, there is provided novel enzymes, as well as active fragments, analogs and derivatives thereof.

In accordance with another aspect of the present invention, there are provided isolated nucleic acid molecules encoding enzymes of the present invention including mRNAs, DNAs, cDNAs, genomic DNAs as well as active analogs and fragments of such enzymes.

In accordance with another aspect of the present invention there are provided isolated nucleic acid molecules encoding mature polypeptides expressed by the DNA in SEQ ID Nos: 1, 3, 5, 7, 9, 11.

In accordance with yet a further aspect of the present invention, there is provided a process for producing such polypeptide by recombinant techniques comprising culturing recombinant prokaryotic and/or eukaryotic host cells, containing a nucleic acid sequence encoding an enzyme of the present invention, under conditions promoting expression of said enzyme and subsequent recovery of said enzyme.

In accordance with yet a further aspect of the present invention, there is provided a process for utilizing such enzymes, or polynucleotide encoding such enzymes for polymerizing DNA.

In accordance with yet a further aspect of the present invention, there is also provided nucleic acid probes comprising nucleic acid molecules of sufficient length to specifically hybridize to a nucleic acid sequence of the present invention.

In accordance with yet a further aspect of the present invention, there is provided a process for utilizing such enzymes, or polynucleotides encoding such enzymes, for *in vitro* purposes related to scientific research, for example, to generate probes for identifying similar sequences which might encode similar enzymes from other organisms.

These and other aspects of the present invention should be apparent to those skilled in the art from the teachings herein.

Brief Description of the Drawings

The following drawings are illustrative of embodiments of the invention and are not meant to limit the scope of the invention as encompassed by the claims. Sequencing was performed using
5 a 378 automated DNA sequencer (Applied Biosystems, Inc.).

FIGURE 1 shows the nucleotide and deduced amino acid sequence of DNA polymerase (3py1) from *Ammonifex degensii*.

10 FIGURE 2 shows the nucleotide and deduced amino acid sequence of DNA polymerase (1PY2) from *Pyrolobus fumarius*.

FIGURE 3 shows the nucleotide and deduced amino acid sequence of DNA polymerase (5PY1) from *Archaeoglobus lithotrophicus*.

15 FIGURE 4 shows the nucleotide and deduced amino acid sequence of DNA polymerase (23PY1) from *Metallosphaera prunae*.

FIGURE 5 shows the nucleotide and deduced amino acid sequence of DNA polymerase
20 (29PY1) from *Desulfurococcus*.

FIGURE 6 shows the nucleotide and deduced amino acid sequence of DNA polymerase (34PY1) from *Aquifex VF-5*.

Description of the Preferred Embodiments

The term "gene" means the segment of DNA involved in producing a polypeptide chain: it includes regions preceding and following the coding region (leader and trailer) as well as intervening sequences (introns) between individual coding segments (exons).

A coding sequence is "operably linked to" another coding sequence when RNA polymerase will transcribe the two coding sequences into a single mRNA, which is then translated into a single polypeptide having amino acids derived from both coding sequences. The coding sequences need not be contiguous to one another so long as the expressed sequences are ultimately processed to produce the desired protein.

"Recombinant" enzymes refer to enzymes produced by recombinant DNA techniques: *i.e.*, produced from cells transformed by an exogenous DNA construct encoding the desired enzyme. "Synthetic" enzymes are those prepared by chemical synthesis.

A DNA "coding sequence of" or a "nucleotide sequence encoding" a particular enzyme, is a DNA sequence which is transcribed and translated into an enzyme when placed under the control of appropriate regulatory sequences. A "promotor sequence" is a DNA regulatory region capable of binding RNA polymerase in a cell and initiating transcription of a downstream (3' direction) coding sequence. The promoter is part of the DNA sequence. This sequence region has a start codon at its 3' terminus. The promoter sequence does include the minimum number of bases where elements necessary to initiate transcription at levels detectable above background. However, after the RNA polymerase binds the sequence and transcription is initiated at the start codon (3' terminus with a promoter), transcription proceeds downstream in the 3' direction. Within the promotor sequence will be found a transcription initiation site (conveniently defined by mapping with nuclease S1) as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase.

The present invention provides purified thermostable enzymes that catalyze DNA synthesis by addition of deoxynucleotides to the 3' end of a polynucleotide chain, using a complementary polynucleotide strand as a template. An exemplary purified enzyme is a polymerase derived from an organism referred herein as "Ammonifex degensii KC4" is a gram negative, chemolithoautotrophic eubacteria and has a very high temperature optimum. Ammonifex degensii KC4 was discovered in a deep sea isolate from the Middle Atlantic Ridge. Ammonifex degensii

KC4 grows optimally at 70°C and pH 7.0 in a low salt medium. This exemplary enzyme is shown in Figure 1.

5 The polynucleotide encoding SEQ ID NO:1 was originally recovered from a genomic gene library derived from *Ammonifex degensii* KC4 as described below. It contains an open reading frame encoding a protein of 867 amino acid residues.

10 In one embodiment, the representative polymerase of SEQ ID NO:1 of the present invention has a molecular weight of about 95.6 kilodaltons as measured by SDS-PAGE gel electrophoresis and an inferred molecular weight from the nucleotide sequence of the gene. This purified enzyme may be used to polymerize DNA where desired. The polymerase enzyme of the present invention has a very high thermostability and has the closest homology to polymerase from *Bacillus stearothermophilus* with 56% identity and 75% similarity at the amino acid level.

15 In accordance with an aspect of the present invention, there are provided isolated nucleic acid molecules (polynucleotides) which encode for the mature enzymes having the deduced amino acid sequence of Figures 1-6 and SEQ ID NOs:1, 3, 5, 7, 9, 11.

This invention, in addition to the isolated nucleic acid molecule encoding an polymerase enzyme disclosed in Figures 1-6 (SEQ ID NOs:1, 3, 5, 7, 9, 11), also provides substantially similar sequences. Isolated nucleic acid sequences are substantially similar if: (i) they are capable of hybridizing under stringent conditions, hereinafter described, to SEQ ID NO:1; or (ii) 20 they encode DNA sequences which are degenerate to SEQ ID NO:1. Degenerate DNA sequences encode the amino acid sequence of SEQ ID NO:2, but have variations in the nucleotide coding sequences. As used herein, "substantially similar" refers to the sequences having similar identity to the sequences of the instant invention. The nucleotide sequences that are substantially similar can be identified by hybridization or by sequence comparison. Enzyme sequences that are 25 substantially similar can be identified by one or more of the following: proteolytic digestion, gel electrophoresis and/or microsequencing. One means for isolating a nucleic acid molecule encoding a polymerase enzyme is to probe a genomic gene library with a natural or artificially designed probe using art recognized procedures (see, for example: Current Protocols in Molecular Biology, Ausubel F.M. *et al.* (EDS.) Green Publishing Company Assoc. and John Wiley Interscience, New York, 1989, 1992). It is appreciated to one skilled in the art that SEQ 30 ID NO:1, or fragments thereof (comprising at least 10 contiguous nucleotides and at least 70%

complementary to a target sequence), is a particularly useful probe. Other particular useful probes for this purpose are hybridizable fragments to the sequences of SEQ ID NO:1 (*i.e.*, comprising at least 10 contiguous nucleotides and at least 70% complementary to a target sequence).

With respect to nucleic acid sequences which hybridize to specific nucleic acid sequences disclosed herein, hybridization may be carried out under conditions of reduced stringency, medium stringency or even stringent conditions. As an example of oligonucleotide hybridization, a polymer membrane containing immobilized denatured nucleic acid is first prehybridized for 30 minutes at 45°C in a solution consisting of 0.9 M NaCl, 50 mM NaH₂PO₄, pH 7.0, 5.0 mM Na₂EDTA, 0.5% SDS, 10X Denhardt's, and 0.5 mg/mL polyriboadenylic acid. Approximately 2×10^7 cpm (specific activity $4-9 \times 10^6$ cpm/ μ g) of ³²P end-labeled oligonucleotide probe are then added to the solution. After 12-16 hours of incubation, the membrane is washed for 30 minutes at room temperature in 1X SET (150 mM NaCl, 20 mM Tris hydrochloride, pH 7.8, 1 mM Na₂EDTA) containing 0.5% SDS, followed by a 30 minute wash in fresh 1X SET at Tm-10°C for the oligo-nucleotide probe. The membrane is then exposed to auto-radiographic film for detection of hybridization signals.

In nucleic acid hybridization reactions, the conditions used to achieve a particular level of stringency will vary, depending on the nature of the nucleic acids being hybridized. For example, the length, degree of complementarity, nucleotide sequence composition (*e.g.*, GC v. AT content), and nucleic acid type (*e.g.*, RNA v. DNA) of the hybridizing regions of the nucleic acids can be considered in selecting hybridization conditions. An additional consideration is whether one of the nucleic acids is immobilized, for example, on a filter.

An example of progressively higher stringency conditions is as follows: $2 \times$ SSC/0.1% SDS at about room temperature (hybridization conditions); $0.2 \times$ SSC/0.1% SDS at about room temperature (low stringency conditions); $0.2 \times$ SSC/0.1% SDS at about 42°C (moderate stringency conditions); and $0.1 \times$ SSC at about 68°C (high stringency conditions). Washing can be carried out using only one of these conditions, *e.g.*, high stringency conditions, or each of the conditions can be used, *e.g.*, for 10-15 minutes each, in the order listed above, repeating any or all of the steps listed. However, as mentioned above, optimal conditions will vary, depending on the particular hybridization reaction involved, and can be determined empirically.

"Identity" as the term is used herein, refers to a polynucleotide sequence which comprises a percentage of the same bases as a reference polynucleotide (SEQ ID NO:1). For example, a polynucleotide which is at least 90% identical to a reference polynucleotide, has polynucleotide bases which are identical in 90% of the bases which make up the reference polynucleotide and may have different bases in 10% of the bases which comprise that polynucleotide sequence.

The present invention also relates to polynucleotides which differ from the reference polynucleotide such that the changes are silent changes, for example the changes do not alter the amino acid sequence encoded by the polynucleotide. The present invention also relates to nucleotide changes which result in amino acid substitutions, additions, deletions, fusions and truncations in the enzyme encoded by the reference polynucleotide (SEQ ID NO:1). In a preferred aspect of the invention these enzymes retain the same biological action as the enzyme encoded by the reference polynucleotide.

It is also appreciated that such probes can be and are preferably labeled with an analytically detectable reagent to facilitate identification of the probe. Useful reagents include but are not limited to radioactivity, fluorescent dyes or enzymes capable of catalyzing the formation of a detectable product. The probes are thus useful to isolate complementary copies of DNA from other animal sources or to screen such sources for related sequences.

The present invention provides substantially pure polymerase enzymes. The term "substantially pure" is used herein to describe a molecule, such as a polypeptide (*e.g.*, a polymerase polypeptide, or a fragment thereof) that is substantially free of other proteins, lipids, carbohydrates, nucleic acids, and other biological materials with which it is naturally associated. For example, a substantially pure molecule, such as a polypeptide, can be at least 60%, by dry weight, the molecule of interest. The purity of the polypeptides can be determined using standard methods including, *e.g.*, polyacrylamide gel electrophoresis (*e.g.*, SDS-PAGE), column chromatography (*e.g.*, high performance liquid chromatography (HPLC)), and amino-terminal amino acid sequence analysis.

Polymerase polypeptides included in the invention can have one of the amino acid sequences of polymerases shown in Figures 1 through 6 (SEQ ID Nos:2, 4, 6, 8, 10, 12), for example, the amino acid sequence of *Ammonifex degensii* KC4 (SEQ ID NO:2). Polymerase

polypeptides, such as those isolated from *Ammonifex degensii* KC4, can be characterized by polymerizing DNA.

Also included in the invention are polypeptides having sequences that are "substantially identical" to the sequence of a polymerase polypeptide, such as one of SEQ ID NO:2, *e.g.*, SEQ ID NO:4. A "substantially identical" amino acid sequence is a sequence that differs from a reference sequence only by conservative amino acid substitutions, for example, substitutions of one amino acid for another of the same class (*e.g.*, substitution of one hydrophobic amino acid, such as isoleucine, valine, leucine, or methionine, for another, or substitution of one polar amino acid for another, such as substitution of arginine for lysine, glutamic acid for aspartic acid, or glutamine for asparagine), or by one or more non-conservative substitutions, deletions, or insertions, provided that the polypeptide retains at least one polymerase-specific activity or a polymerase-specific epitope. For example, one or more amino acids can be deleted from a polymerase polypeptide, resulting in modification of the structure of the polypeptide, without significantly altering its biological activity. For example, amino- or carboxyl-terminal amino acids that are not required for polymerase biological activity, can be removed. Such modifications can result in the development of smaller active polymerase polypeptides.

Other polymerase polypeptides included in the invention are polypeptides having amino acid sequences that are at least 50% identical to the amino acid sequence of a polymerase polypeptide, such as any of polymerases in SEQ ID Nos:2, 4, 6, 8, 10, 12, *e.g.*, SEQ ID NO:12. The length of comparison in determining amino acid sequence homology can be, for example, at least 15 amino acids, for example, at least 20, 25, or 35 amino acids. Homology can be measured using standard sequence analysis software (*e.g.*, Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705; also see Ausubel, *et al.*, *supra*). Such procedures and algorithms include, for example, a BLAST program (Basic Local Alignment Search Tool at the National Center for Biological Information), ALIGN, AMAS (Analysis of Multiply Aligned Sequences), AMPS (Protein Multiple Sequence Alignment), ASSET (Aligned Segment Statistical Evaluation Tool), BANDS, BESTSCOR, BIOSCAN (Biological Sequence Comparative Analysis Node), BLIMPS (BLOCKS IMPROVED Searcher), FASTA, Intervals & Points, BMB, CLUSTAL V, CLUSTAL W, CONSENSUS, LCONSENSUS, WCONSENSUS, Smith-Waterman algorithm.

DARWIN, Las Vegas algorithm, FNAT (Forced Nucleotide Alignment Tool), Framcalign, Framesearch, DYNAMIC, FILTER, FSAP (Fristensky Sequence Analysis Package), GAP (Global Alignment Program), GENAL, GIBBS, GenQuest, ISSC (Sensitive Sequence Comparison),
5 LALIGN (Local Sequence Alignment), LCP (Local Content Program), MACAW (Multiple Alignment Construction & Analysis Workbench), MAP (Multiple Alignment Program), MBLKP, MBLKN, PIMA (Pattern-Induced Multi-sequence Alignment), SAGA (Sequence Alignment by Genetic ALgorithm) and WHAT-IF.

The invention also includes fragments of polymerase polypeptides that retain at least one
10 polymerase-specific activity or epitope. Polymerase activity can be assayed by examining the polymerizing of DNA. For example, a polymerase polypeptide fragment containing, *e.g.*, at least 8-10 amino acids can be used as an immunogen in the production of polymerase-specific antibodies. The fragment can contain, for example, an amino acid sequence that is conserved in polymerases, and this amino acid sequence can contain amino acids that are conserved in
15 polymerases. Such fragments can easily be identified by comparing the sequences of polymerases found in Figures 1-X. In addition to their use as peptide immunogens, the above-described polymerase fragments can be used in immunoassays, such as ELISAs, to detect the presence of polymerase-specific antibodies in samples.

The polymerase polypeptides of the invention can be obtained using any of several
20 standard methods. For example, polymerase polypeptides can be produced in a standard recombinant expression systems (see below), chemically synthesized (this approach may be limited to small polymerase peptide fragments), or purified from organisms in which they are naturally expressed.

The invention also provides isolated nucleic acid molecules that encode the polymerase
25 polypeptides described above, as well as fragments thereof. For example, nucleic acids that encode any of SEQ ID NOs:1, 3, 5, 7, 9, 11 are included in the invention. These nucleic acids can contain naturally occurring nucleotide sequences, or sequences that differ from those of the naturally occurring nucleic acids that encode polymerases, but encode the same amino acids, due to the degeneracy of the genetic code. The nucleic acids of the invention can contain DNA or
30 RNA nucleotides, or combinations or modifications thereof. Exemplary nucleic acids of the invention are shown in SEQ ID NO:1.

By "isolated nucleic acid" is meant a nucleic acid, *e.g.*, a DNA or RNA molecule, that is not immediately contiguous with the 5' and 3' flanking sequences with which it normally is immediately contiguous when present in the naturally occurring genome of the organism from which it is derived. The term thus describes, for example, a nucleic acid that is incorporated into a vector, such as a plasmid or viral vector; a nucleic acid that is incorporated into the genome of a heterologous cell (or the genome of a homologous cell, but at a site different from that at which it naturally occurs); and a nucleic acid that exists as a separate molecule, *e.g.*, a DNA fragment produced by PCR amplification or restriction enzyme digestion, or an RNA molecule produced by *in vitro* transcription. The term also describes a recombinant nucleic acid that forms part of a hybrid gene encoding additional polypeptide sequences that can be used, for example, in the production of a fusion protein.

The nucleic acid molecules of the invention can be used as templates in standard methods for production of polymerase gene products (*e.g.*, polymerase RNAs and polymerase polypeptides). In addition, the nucleic acid molecules that encode polymerase polypeptides (and fragments thereof) and related nucleic acids, such as (1) nucleic acids containing sequences that are complementary to, or that hybridize to, nucleic acids encoding polymerase polypeptides, or fragments thereof (*e.g.*, fragments containing at least 10, 12, 15, 20, or 25 nucleotides); and (2) nucleic acids containing sequences that hybridize to sequences that are complementary to nucleic acids encoding polymerase polypeptides, or fragments thereof (*e.g.*, fragments containing at least 10, 12, 15, 20, or 25 nucleotides); can be used in methods focused on their hybridization properties. For example, as is described in further detail below, such nucleic acid molecules can be used in the following methods: PCR methods for synthesizing polymerase nucleic acids, methods for detecting the presence of an polymerase nucleic acid in a sample, screening methods for identifying nucleic acids encoding new polymerase family members. Oligonucleotide probes useful for screening methods are from 10 to about 150 nucleotides in length. Further, such probes are preferably 10 to about 100 nucleotides in length and more preferably from 10 to about 50 nucleotides in length.

The invention also includes methods for identifying nucleic acid molecules that encode members of the polymerase polypeptide family in addition to SEQ ID NOs:1, 3, 5, 7, 9, 11. In these methods, a sample, *e.g.*, a nucleic acid library, such as a cDNA library, that contains a

nucleic acid encoding a polymerase polypeptide is screened with a polymerase-specific probe, *e.g.*, a polymerase-specific nucleic acid probe. Polymerase-specific nucleic acid probes are nucleic acid molecules (*e.g.*, molecules containing DNA or RNA nucleotides, or combinations or modifications thereof) that specifically hybridize to nucleic acids encoding polymerase polypeptides, or to complementary sequences thereof. The term "polymerase-specific probe", in the context of this method of invention, refers to probes that bind to nucleic acids encoding polymerase polypeptides, or to complementary sequences thereof, to a detectably greater extent than to nucleic acids encoding other enzymes, or to complementary sequences thereof.

The invention facilitates production of polymerase-specific nucleic acid probes. Methods for obtaining such probes can be designed based on the amino acid sequences shown in Figure 1. The probes, which can contain at least 10, *e.g.*, 15, 25, 35, 50, 100, or 150 nucleotides, can be produced using any of several standard methods (see, *e.g.*, Ausubel, *et al.*, *supra*). For example, preferably, the probes are generated using PCR amplification methods. In these methods, primers are designed that correspond to polymerase-conserved sequences (see Figure 1), which can include polymerase-specific amino acids, and the resulting PCR product is used as a probe to screen a nucleic acid library, such as a cDNA library.

The coding sequences for the polymerase enzymes of the present invention were identified by preparing an *Ammonifex degensii* KC4 genomic DNA library, for example, and screening the library for the clones having polymerase activity. Such methods for constructing a genomic gene library are well-known in the art. One means, for example, comprises shearing DNA isolated from *Ammonifex degensii* KC4 by physical disruption. A small amount of the sheared DNA is checked on an agarose gel to verify that the majority of the DNA is in the desired size range (approximately 3-6 kb). The DNA is then blunt ended using Mung Bean Nuclease, incubated at 37°C and phenol/chloroform extracted. The DNA is then methylated using Eco RI Methylase. Eco RI linkers are then ligated to the blunt ends through the use of T4 DNA ligase and incubation at 4°C. The ligation reaction is then terminated and the DNA is cut-back with Eco RI restriction enzyme. The DNA is then size fractionated on a sucrose gradient following procedures known in the art, for example, Maniatis, T., *et al.*, Molecular Cloning, Cold Spring Harbor Press, New York, 1982, which is hereby incorporated by reference in its entirety.

A plate assay is then performed to get an approximate concentration of the DNA. Ligation reactions are then performed and 1 μ l of the ligation reaction is packaged to construct a library. Packaging, for example, may occur through the use of purified (λ gt11 phage arms cut with EcoRI and DNA cut with EcoRI after attaching EcoRI linkers. The DNA and (λ gt11 arms are ligated with DNA ligase. The ligated DNA is then packaged into infectious phage particles. The packaged phages are used to infect *E. coli* cultures and the infected cells are spread on agar plates to yield plates carrying thousands of individual phage plaques. The library is then amplified.

Fragments of the full length gene of the present invention may be used as a hybridization probe for a cDNA or a genomic library to isolate the full length DNA and to isolate other DNAs which have a high sequence similarity to the gene or similar biological activity. Probes of this type have at least 10, preferably at least 15, and even more preferably at least 30 bases and may contain, for example, at least 50 or more bases. The probe may also be used to identify a DNA clone corresponding to a full length transcript and a genomic clone or clones that contain the complete gene including regulatory and promotor regions, exons, and introns.

The isolated nucleic acid sequences and other enzymes may then be measured for retention of biological activity characteristic to the enzyme of the present invention, for example, in an assay for detecting enzymatic polymerase activity. Such enzymes include truncated forms of polymerase, and variants such as deletion and insertion variants.

The polynucleotide of the present invention may be in the form of DNA which DNA includes cDNA, genomic DNA, and synthetic DNA. The DNA may be double-stranded or single-stranded, and if single stranded may be the coding strand or non-coding (anti-sense) strand. The coding sequence which encodes the mature enzyme may be identical to the coding sequences shown in Figures 1-6, or may be a different coding sequence which coding sequence, as a result of the redundancy or degeneracy of the genetic code, encodes the same mature enzyme as the DNA of Figures 1-6 (*e.g.*, SEQ ID NO:1).

The polynucleotide which encodes the mature enzyme of Figure 1 (*e.g.*, SEQ ID NO:1) may include, but is not limited to: only the coding sequence for the mature enzyme; the coding sequence for the mature enzyme and additional coding sequence such as a leader sequence or a

proprotein sequence: the coding sequence for the mature enzyme (and optionally additional coding sequence) and non-coding sequence, such as introns or non-coding sequence 5' and/or 3' of the coding sequence for the mature enzyme.

5 Thus, the term "polynucleotide encoding an enzyme (protein)" encompasses a polynucleotide which includes only coding sequence for the enzyme as well as a polynucleotide which includes additional coding and/or non-coding sequence.

The present invention further relates to variants of the hereinabove described polynucleotides which encode for fragments, analogs and derivatives of the enzyme having the deduced amino acid sequence of Figure 1 (*e.g.*, SEQ ID NO:2). The variant of the
10 polynucleotide may be a naturally occurring allelic variant of the polynucleotide or a non-naturally occurring variant of the polynucleotide.

Thus, the present invention includes polynucleotides encoding the same mature enzyme as shown in Figure 1 as well as variants of such polynucleotides which variants encode for a fragment, derivative or analog of the enzyme of Figure 1. Such nucleotide variants include
15 deletion variants, substitution variants and addition or insertion variants.

As hereinabove indicated, the polynucleotide may have a coding sequence which is a naturally occurring allelic variant of the coding sequence shown in Figure 1. As known in the art, an allelic variant is an alternate form of a polynucleotide sequence which may have a
20 substitution, deletion or addition of one or more nucleotides, which does not substantially alter the function of the encoded enzyme.

The present invention also includes polynucleotides, wherein the coding sequence for the mature enzyme may be fused in the same reading frame to a polynucleotide sequence which aids in expression and secretion of an enzyme from a host cell, for example, a leader sequence which
25 functions to control transport of an enzyme from the cell. The enzyme having a leader sequence is a preprotein and may have the leader sequence cleaved by the host cell to form the mature form of the enzyme. The polynucleotides may also encode for a preprotein which is the mature protein plus additional 5' amino acid residues. A mature protein having a prosequence is a preprotein and is an inactive form of the protein. Once the prosequence is cleaved an active
30 mature protein remains.

Thus, for example, the polynucleotide of the present invention may encode for a mature enzyme, or for an enzyme having a prosequence or for an enzyme having both a prosequence and a presequence (leader sequence).

5 The present invention further relates to polynucleotides which hybridize to the hereinabove-described sequences if there is at least 70%, preferably at least 90%, and more preferably at least 95% identity between the sequences. The present invention particularly relates to polynucleotides which hybridize under stringent conditions to the hereinabove-described polynucleotides. As herein used, the term "stringent conditions" means hybridization will occur
10 only if there is at least 95% and preferably at least 97% identity between the sequences. The polynucleotides which hybridize to the hereinabove described polynucleotides in a preferred embodiment encode enzymes which either retain substantially the same biological function or activity as the mature enzyme encoded by the DNA of Figure 1.

Alternatively, the polynucleotide may have at least 15 bases, preferably at least 30 bases,
15 and more preferably at least 50 bases which hybridize to a polynucleotide of the present invention and which has an identity thereto, as hereinabove described, and which may or may not retain activity. For example, such polynucleotides may be employed as probes for the polynucleotide of SEQ ID NO:1, for example, for recovery of the polynucleotide or as a PCR primer.

Thus, the present invention is directed to polynucleotides having at least a 70% identity,
20 preferably at least 90% identity and more preferably at least a 95% identity to a polynucleotide which encodes the enzyme of SEQ ID NO:1 as well as fragments thereof, which fragments have at least 30 bases and preferably at least 50 bases to enzymes encoded by such polynucleotides.

The present invention further relates to an enzyme which has the deduced amino acid sequence of Figures 1-6, as well as fragments, analogs and derivatives of such enzyme.

25 The terms "fragment," "derivative" and "analog" when referring to the enzyme of Figure 1 means an enzyme which retains essentially the same biological function or activity as such enzyme. Thus, an analog includes a proprotein which can be activated by cleavage of the proprotein portion to produce an active mature enzyme.

The enzyme of the present invention may be a recombinant enzyme, a natural enzyme or
30 a synthetic enzyme, preferably a recombinant enzyme.

The fragment, derivative or analog of the enzyme of Figure 1 may be (i) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue
5 may or may not be one encoded by the genetic code, or (ii) one in which one or more of the amino acid residues includes a substituent group, or (iii) one in which the mature enzyme is fused with another compound, such as a compound to increase the half-life of the enzyme (for example, polyethylene glycol), or (iv) one in which the additional amino acids are fused to the mature enzyme, such as a leader or secretory sequence or a sequence which is employed for purification
10 of the mature enzyme or a proprotein sequence. Such fragments, derivatives and analogs are deemed to be within the scope of those skilled in the art from the teachings herein.

The enzymes and polynucleotides of the present invention are preferably provided in an isolated form, and preferably are purified to homogeneity.

The term "isolated" means that the material is removed from its original environment (e.g.,
15 the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or enzyme present in a living animal is not isolated, but the same polynucleotide or enzyme, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotides could be part of a vector and/or such polynucleotides or enzymes could be part of a composition, and still be isolated in that such vector or composition is not part
20 of its natural environment.

The enzymes of the present invention include an enzyme of Figures 1-6 (in particular the mature enzyme) as well as enzymes which have at least 70% similarity (preferably at least 70% identity) to an enzyme of Figures 1-6 and more preferably at least 90% similarity (more preferably at least 90% identity) to an enzymes of Figures 1-6 and still more preferably at least
25 95% similarity (still more preferably at least 95% identity) to an enzyme of Figures 1-6 and also include portions of such enzymes with such portion of the enzyme generally containing at least 30 amino acids and more preferably at least 50 amino acids.

As known in the art "similarity" between two enzymes is determined by comparing the amino acid sequence and its conserved amino acid substitutes of one enzyme to the sequence of
30 a second enzyme. Similarity in nucleic acid and amino acid sequences may be determined by procedures and algorithms which are well-known in the art. Such procedures and algorithms

include, for example, a BLAST program (Basic Local Alignment Search Tool at the National Center for Biological Information), ALIGN, AMAS (Analysis of Multiply Aligned Sequences), AMPS (Protein Multiple Sequence Alignment), ASSET (Aligned Segment Statistical Evaluation Tool), BANDS, BESTSCOR, BIOSCAN (Biological Sequence Comparative Analysis Node), BLIMPS (BLOCKS IMPROVED Searcher), FASTA, Intervals & Points, BMB, CLUSTAL V, CLUSTAL W, CONSENSUS, LCONSENSUS, WCONSENSUS, Smith-Waterman algorithm, DARWIN, Las Vegas algorithm, FNAT (Forced Nucleotide Alignment Tool), Framealign, Framesearch, DYNAMIC, FILTER, FSAP (Fristensky Sequence Analysis Package), GAP (Global Alignment Program), GENAL, GIBBS, GenQuest, ISSC (Sensitive Sequence Comparison), LALIGN (Local Sequence Alignment), LCP (Local Content Program), MACAW (Multiple Alignment Construction & Analysis Workbench), MAP (Multiple Alignment Program), MBLKP, MBLKN, PIMA (Pattern-Induced Multi-sequence Alignment), SAGA (Sequence Alignment by Genetic Algorithm) and WHAT-IF.

A variant, *i.e.* a "fragment", "analog" or "derivative" enzyme, and reference enzyme may differ in amino acid sequence by one or more substitutions, additions, deletions, fusions and truncations, which may be present in any combination.

Among preferred variants are those that vary from a reference by conservative amino acid substitutions. Such substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe, Tyr.

Most highly preferred are variants which retain the same biological function and activity as the reference polypeptide from which it varies.

Fragments or portions of the enzymes of the present invention may be employed for producing the corresponding full-length enzyme by peptide synthesis; therefore, the fragments may be employed as intermediates for producing the full-length enzymes. Fragments or portions of the polynucleotides of the present invention may be used to synthesize full-length polynucleotides of the present invention.

The present invention also relates to vectors which include polynucleotides of the present invention, host cells which are genetically engineered with vectors of the invention and the production of enzymes of the invention by recombinant techniques.

5 Host cells are genetically engineered (transduced or transformed or transfected) with the vectors containing the polynucleotides of this invention. Such vectors may be, for example, a cloning vector or an expression vector. The vector may be, for example, in the form of a plasmid, a viral particle, a phage, etc. The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or
10 amplifying the genes of the present invention. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to the ordinarily skilled artisan.

The polynucleotides of the present invention may be employed for producing enzymes by recombinant techniques. Thus, for example, the polynucleotide may be included in any one of
15 a variety of expression vectors for expressing an enzyme. Such vectors include chromosomal, nonchromosomal and synthetic DNA sequences, e.g., derivatives of SV40; bacterial plasmids; phage DNA; baculovirus; yeast plasmids; vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, and pseudorabies. However, any other vector may be used as long as it is replicable and viable in the host.

20 The appropriate DNA sequence may be inserted into the vector by a variety of procedures. In general, the DNA sequence is inserted into an appropriate restriction endonuclease site(s) by procedures known in the art. Such procedures and others are deemed to be within the scope of those skilled in the art.

The DNA sequence in the expression vector is operatively linked to an appropriate
25 expression control sequence(s) (promoter) to direct mRNA synthesis. As representative examples of such promoters, there may be mentioned: LTR or SV40 promoter, the *E. coli*, *lac* or *trp*, the phage lambda P_L promoter and other promoters known to control expression of genes in prokaryotic or eukaryotic cells or their viruses. The expression vector also contains a ribosome binding site for translation initiation and a transcription terminator. The vector may also include
30 appropriate sequences for amplifying expression.

In addition, the expression vectors preferably contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in *E. coli*.

The vector containing the appropriate DNA sequence as hereinabove described, as well as an appropriate promoter or control sequence, may be employed to transform an appropriate host to permit the host to express the protein.

As representative examples of appropriate hosts, there may be mentioned: bacterial cells, such as *E. coli*, *Streptomyces*, *Bacillus subtilis*; fungal cells, such as yeast; insect cells such as *Drosophila S2* and *Spodoptera Sf9*; animal cells such as CHO, COS or Bowes melanoma; adenoviruses; plant cells, etc. The selection of an appropriate host is deemed to be within the scope of those skilled in the art from the teachings herein.

More particularly, the present invention also includes recombinant constructs comprising one or more of the sequences as broadly described above. The constructs comprise a vector, such as a plasmid or viral vector, into which a sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. The following vectors are provided by way of example; Bacterial: pQE70, pQE60, pQE-9 (Qiagen), pBluescript II (Stratagene); pTRC99a, pKK223-3, pDR540, pRIT2T (Pharmacia); Eukaryotic: pXT1, pSG5 (Stratagene) pSVK3, pBPV, pMSG, pSVLSV40 (Pharmacia). However, any other plasmid or vector may be used as long as they are replicable and viable in the host.

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda P_R, P_L and trp. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

In a further embodiment, the present invention relates to host cells containing the above-described constructs. The host cell can be a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-Dextran mediated transfection, or electroporation (Davis, L., Dibner, M., Battey, I., Basic Methods in Molecular Biology, (1986)).

The constructs in host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence. Alternatively, the enzymes of the invention can be synthetically produced by conventional peptide synthesizers.

Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, N.Y., (1989), the disclosure of which is hereby incorporated by reference.

Transcription of the DNA encoding the enzymes of the present invention by higher eukaryotes is increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp that act on a promoter to increase its transcription. Examples include the SV40 enhancer on the late side of the replication origin bp 100 to 270, a cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, e.g., the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), α -factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated enzyme. Optionally, the

heterologous sequence can encode a fusion enzyme including an N-terminal identification peptide imparting desired characteristics. e.g., stabilization or simplified purification of expressed recombinant product.

5 Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for
10 transformation include *E. coli*, *Bacillus subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*, although others may also be employed as a matter of choice.

As a representative but nonlimiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially
15 available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM1 (Promega Biotec, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed.

Following transformation of a suitable host strain and growth of the host strain to an
20 appropriate cell density, the selected promoter is induced by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period.

Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Microbial cells employed in expression of proteins can be disrupted by any convenient
25 method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, such methods are well known to those skilled in the art.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, *Cell*, 23:175 (1981), and other cell lines capable of expressing
30 a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and

also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

The enzyme can be recovered and purified from recombinant cell cultures by methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

The enzymes of the present invention may be a naturally purified product, or a product of chemical synthetic procedures, or produced by recombinant techniques from a prokaryotic or eukaryotic host (for example, by bacterial, yeast, higher plant, insect and mammalian cells in culture). Depending upon the host employed in a recombinant production procedure, the enzymes of the present invention may be glycosylated or may be non-glycosylated. Enzymes of the invention may or may not also include an initial methionine amino acid residue.

The enzyme of this invention may be employed for any purpose in which such enzyme activity is necessary or desired. In a preferred embodiment the enzyme is employed for catalyzing DNA synthesis by addition of deoxynucleotides to the 3' end of a polynucleotide chain, using a complementary polynucleotide strand as a template.

In a preferred embodiment, the enzyme of the present invention is a thermostable enzyme which is stable to heat and is heat resistant and polymerizes DNA, *i.e.*, the enzyme is able to renature and regain activity after a brief (*i.e.*, 5 to 30 seconds), or longer period, for example, minutes or hours, exposure to temperatures of up to 70°C and has a temperature optimum above 60°C.

The enzymes, their fragments or other derivatives, or analogs thereof, or cells expressing them can be used as an immunogen to produce antibodies thereto. These antibodies can be, for example, polyclonal or monoclonal antibodies. The present invention also includes chimeric, single chain, and humanized antibodies, as well as Fab fragments, or the product of an Fab

expression library. Various procedures known in the art may be used for the production of such antibodies and fragments.

Antibodies generated against the enzymes corresponding to a sequence of the present invention can be obtained by direct injection of the enzymes into an animal or by administering the enzymes to an animal, preferably a nonhuman. The antibody so obtained will then bind the enzymes itself. In this manner, even a sequence encoding only a fragment of the enzymes can be used to generate antibodies binding the whole native enzymes. Such antibodies can then be used to isolate the enzyme from cells expressing that enzyme.

For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler and Milstein, 1975, *Nature*, 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor et al., 1983, *Immunology Today* 4:72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole, et al., 1985, in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96).

Techniques described for the production of single chain antibodies (U.S. Patent 4,946,778) can be adapted to produce single chain antibodies to immunogenic enzyme products of this invention. Also, transgenic mice may be used to express humanized antibodies to immunogenic enzyme products of this invention.

Antibodies generated against the enzyme of the present invention may be used in screening for similar enzymes from other organisms and samples. Such screening techniques are known in the art, for example, one such screening assay is described in "Methods for Measuring Cellulase Activities", *Methods in Enzymology*, Vol 160, pp. 87-116, which is hereby incorporated by reference in its entirety. Antibodies may also be employed as a probe to screen gene libraries generated from this or other organisms to identify this or cross reactive activities.

Isolation and purification of polypeptides produced in the systems described above can be carried out using conventional methods, appropriate for the particular system. For example, preparative chromatography and immunological separations employing antibodies, such as monoclonal or polyclonal antibodies, can be used.

The term "antibody," as used herein, refers to intact immunoglobulin molecules, as well as fragments of immunoglobulin molecules, such as Fab, Fab', (Fab')₂, Fv, and SCA fragments.

that are capable of binding to an epitope of a polymerase polypeptide. These antibody fragments, which retain some ability to selectively bind to the antigen (*e.g.*, a polymerase antigen) of the antibody from which they are derived, can be made using well known methods in the art (see, *e.g.*, Harlow and Lane, *supra*), and are described further, as follows.

(1) A Fab fragment consists of a monovalent antigen-binding fragment of an antibody molecule, and can be produced by digestion of a whole antibody molecule with the enzyme papain, to yield a fragment consisting of an intact light chain and a portion of a heavy chain.

(2) A Fab' fragment of an antibody molecule can be obtained by treating a whole antibody molecule with pepsin, followed by reduction, to yield a molecule consisting of an intact light chain and a portion of a heavy chain. Two Fab' fragments are obtained per antibody molecule treated in this manner.

(3) A (Fab')₂ fragment of an antibody can be obtained by treating a whole antibody molecule with the enzyme pepsin, without subsequent reduction. A (Fab')₂ fragment is a dimer of two Fab' fragments, held together by two disulfide bonds.

(4) An Fv fragment is defined as a genetically engineered fragment containing the variable region of a light chain and the variable region of a heavy chain expressed as two chains.

(5) A single chain antibody ("SCA") is a genetically engineered single chain molecule containing the variable region of a light chain and the variable region of a heavy chain, linked by a suitable, flexible polypeptide linker.

As used in this invention, the term "epitope" refers to an antigenic determinant on an antigen, such as a polymerase polypeptide, to which the paratope of an antibody, such as a polymerase-specific antibody, binds. Antigenic determinants usually consist of chemically active surface groupings of molecules, such as amino acids or sugar side chains, and can have specific three-dimensional structural characteristics, as well as specific charge characteristics.

As is mentioned above, antigens that can be used in producing polymerase-specific antibodies include polymerase polypeptides, *e.g.*, any of the polymerases shown in Figures 1-X polypeptide fragments. The polypeptide or peptide used to immunize an animal can be obtained by standard recombinant, chemical synthetic, or purification methods. As is well known in the

art, in order to increase immunogenicity, an antigen can be conjugated to a carrier protein. Commonly used carriers include keyhole limpet hemocyanin (KLH), thyroglobulin, bovine serum albumin (BSA), and tetanus toxoid. The coupled peptide is then used to immunize the animal (e.g., a mouse, a rat, or a rabbit). In addition to such carriers, well known adjuvants can be administered with the antigen to facilitate induction of a strong immune response.

Polymerase-specific polyclonal and monoclonal antibodies can be purified, for example, by binding to, and elution from, a matrix containing a polymerase polypeptide, e.g., the polymerase polypeptide (or fragment thereof) to which the antibodies were raised. Additional methods for antibody purification and concentration are well known in the art and can be practiced with the polymerase-specific antibodies of the invention (see, for example, Coligan, *et al.*, Unit 9, *Current Protocols in Immunology*, Wiley Interscience, 1994).

Anti-idiotypic antibodies corresponding to polymerase-specific antigens are also included in the invention, and can be produced using standard methods. These antibodies are raised to polymerase-specific antibodies, and thus mimic polymerase-specific epitopes.

The members of a pair of molecules (e.g., an antibody-antigen pair or a nucleic acid pair) are said to "specifically bind" to each other if they bind to each other with greater affinity than to other, non-specific molecules. For example, an antibody raised against an antigen to which it binds more efficiently than to a non-specific protein can be described as specifically binding to the antigen. (Similarly, a nucleic acid probe can be described as specifically binding to a nucleic acid target if it forms a specific duplex with the target by base pairing interactions (see above).)

The present invention is further described with reference to the following examples; however, it is to be understood that the present invention is not limited to such examples. All parts or amounts, unless otherwise specified, are by weight.

In one aspect of the invention, a method for producing a polymerase enzyme, such as those shown in Figures 1-6, is provided. The method includes growing a host cell which contains a polynucleotide encoding the enzyme (e.g., SEQ ID Nos:2, 4, 6, 8, 10, 12), under conditions which allow the expression of the nucleic acid, and isolating the enzyme encoded by the nucleic acid. Methods of culturing the host cell are described in the Examples and are known by those of skill in the art.

In order to facilitate understanding of the following examples certain frequently occurring methods and/or terms will be described.

"Plasmids" are designated by a lower case p preceded and/or followed by capital letters and/or numbers. The starting plasmids herein are either commercially available, publicly available on an unrestricted basis, or can be constructed from available plasmids in accord with published procedures. In addition, equivalent plasmids to those described are known in the art and will be apparent to the ordinarily skilled artisan.

"Digestion" of DNA refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes used herein are commercially available and their reaction conditions, cofactors and other requirements were used as would be known to the ordinarily skilled artisan. For analytical purposes, typically 1 μ g of plasmid or DNA fragment is used with about 2 units of enzyme in about 20 μ l of buffer solution. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 μ g of DNA are digested with 20 to 250 units of enzyme in a larger volume. Appropriate buffers and substrate amounts for particular restriction enzymes are specified by the manufacturer. Incubation times of about 1 hour at 37°C are ordinarily used, but may vary in accordance with the supplier's instructions. After digestion the reaction is electrophoresed directly on a polyacrylamide gel to isolate the desired fragment.

Size separation of the cleaved fragments is generally performed using 8 percent polyacrylamide gel described by Goeddel, D. *et al.*, Nucleic Acids Res., 8:4057 (1980), for example.

"Oligonucleotides" refers to either a single stranded polydeoxynucleotide or two complementary polydeoxynucleotide strands which may be chemically synthesized. Such synthetic oligonucleotides may or may not have a 5' phosphate. Those that do not will not ligate to another oligonucleotide without adding a phosphate with an ATP in the presence of a kinase. A synthetic oligonucleotide will ligate to a fragment that has not been dephosphorylated.

"Ligation" refers to the process of forming phosphodiester bonds between two double stranded nucleic acid fragments (Maniatis, T., *et al.*, Id., p. 146). Unless otherwise provided, ligation may be accomplished using known buffers and conditions with 10 units of T4 DNA

ligase ("ligase") per 0.5 µg of approximately equimolar amounts of the DNA fragments to be ligated.

5 Unless otherwise stated, transformation was performed as described in the method of Sambrook, Fritsch and Maniatis, 1989. The following examples are intended to illustrate, but not to limit, the invention. While the procedures described in the examples are typical of those that can be used to carry out certain aspects of the invention, other procedures known to those skilled in the art can also be used. The following materials and methods were used in carrying out the experiments described in the examples.

Example 1

DNA Isolation and Library Construction

The following outlines the procedures used to generate a gene library from a sample.

5

Isolate DNA.

IsoQuick Procedure as per manufacturer's instructions (Orca, Research Inc., Bothell, WA).

Shear DNA

10

Vigorously push and pull DNA through a 25G double-hub needle and 1-cc syringes about 500 times.

Check a small amount (0.5 µg) on a 0.8% agarose gel to make sure the majority of the DNA is in the desired size range (about 3-6 kb).

15

Blunt DNA

Add:

H ₂ O	to a final volume of 405 µl
45 µl	10X Mung Bean Buffer
2.0 µl	Mung Bean Nuclease (150 u/µl)

20

Incubate 37°C, 15 minutes.

Phenol/chloroform extract once.

Chloroform extract once.

Add 1 ml ice cold ethanol to precipitate.

Place on ice for 10 minutes.

25

Spin in microfuge, high speed, 30 minutes.

Wash with 1 ml 70% ethanol.

Spin in microfuge, high speed, 10 minutes and dry.

Methylate DNA

Gently resuspend DNA in 26 μ l TE.

Add:

5 4.0 μ l 10X *EcoR* I Methylase Buffer

0.5 μ l SAM (32 mM)

5.0 μ l *EcoR* I Methylase (40 u/ μ l)

Incubate 37°, 1 hour.

Insure Blunt Ends

10 Add to the methylation reaction:

5.0 μ l 100 mM MgCl₂

8.0 μ l dNTP mix (2.5 mM of each dGTP, dATP, dTTP, dCTP)

4.0 μ l Klenow (5 u/ μ l)

Incubate 12°C, 30 minutes.

15 Add 450 μ l 1X STE.

Phenol/chloroform extract once.

Chloroform extract once.

Add 1 ml ice cold ethanol to precipitate and place on ice for 10 minutes.

Spin in microfuge, high speed, 30 minutes.

20 Wash with 1 ml 70% ethanol.

Spin in microfuge, high speed, 10 minutes and dry.

Adaptor Ligation

Gently resuspend DNA in 8 μ l *EcoR* I adaptors (from Stratagene's cDNA Synthesis Kit).

25 Add:

1.0 μ l 10X Ligation Buffer

1.0 μ l 10 mM rATP

1.0 μ l T4 DNA Ligase (4 u/ μ l)

Incubate 4°C, 2 days.

30

Phosphorylate Adaptors

Heat kill ligation reaction 70°C, 30 minutes.

Add:

5 1.0 µl 10X Ligation Buffer
 2.0 µl 10mM rATP
 6.0 µl H₂O
 1.0 µl Polynucleotide kinase (PNK)

Incubate 37°C, 30 minutes.

10 Add 31 µl H₂O and 5 µl 10X STE.

Size fractionate on a Sephacryl S-500 spin column (pool fractions 1-3).

Phenol/chloroform extract once.

Chloroform extract once.

Add ice cold ethanol to precipitate.

15 Place on ice, 10 minutes.

Spin in microfuge, high speed. 30 minutes.

Wash with 1 ml 70% ethanol.

Spin in microfuge, high speed. 10 minutes and dry.

Resuspend in 10.5 µl TE buffer.

20 Do not plate assay. Instead, ligate directly to arms as above except use 2.5 µl of
DNA and no water.

Sucrose Gradient (2.2 ml) Size Fractionation

Heat sample to 65°C, 10 minutes.

25 Gently load on 2.2 ml sucrose gradient.

Spin in mini-ultracentrifuge, 45K, 20°C. 4 hours (no brake).

Collect fractions by puncturing the bottom of the gradient tube with a 20G needle and
allowing the sucrose to flow through the needle. Collect the first 20 drops in a Falcon
2059 tube then collect 10 1-drop fractions (labelled 1-10). Each drop is about 60 µl in
30 volume.

Run 5 µl of each fraction on a 0.8% agarose gel to check the size.

Pool fractions 1-4 (about 10-1.5 kb) and, in a separate tube, pool fractions 5-7 (about 5-0.5 kb).

Add 1 ml ice cold ethanol to precipitate and place on ice for 10 minutes.

5 Spin in microfuge, high speed, 30 minutes.

Wash with 1 ml 70% ethanol.

Spin in microfuge, high speed, 10 minutes and dry.

Resuspend each in 10 μ l TE buffer.

10 Test Ligation to Lambda Arms

Plate assay to get an approximate concentration. Spot 0.5 μ l of the sample on agarose containing ethidium bromide along with standards (DNA samples of known concentration). View in UV light and estimate concentration compared to the standards. Fraction 1-4 = >1.0 μ g/ μ l. Fraction 5-7 = 500 ng/ μ l.

15 Prepare the following ligation reactions (5 μ l reactions) and incubate 4°C. overnight:

Sample	H ₂ O	10X Ligase Buffer	10mM rATP	Lambd a arms (ZAP)	Insert DNA	T4 DNA Ligase (4 u/ μ l)
Fraction 1-4	0.5 μ l	0.5 μ l	0.5 μ l	1.0 μ l	2.0 μ l	0.5 μ l
Fraction 5-7	0.5 μ l	0.5 μ l	0.5 μ l	1.0 μ l	2.0 μ l	0.5 μ l

20

Test Package and Plate

Package the ligation reactions following manufacturer's protocol.

Stop packaging reactions with 500 μ l SM buffer and pool packaging that came from the same ligation.

25 Titer 1.0 μ l of each on appropriate host (OD₆₀₀ = 1.0) [XLI-Blue MRF]

Add 200 μ l host (in mM MgSO₄) to Falcon 2059 tubes

Inoculate with 1 μ l packaged phage

Incubate 37°C. 15 minutes

Add about 3 ml 48°C top agar

30 [50ml stock containing 150 μ l IPTG (0.5M) and 300 μ l X-GAL (350 mg/ml)]

Plate on 100mm plates and incubate 37°C. overnight.

Amplification of Libraries (5.0×10^5 recombinants from each library)

Add 3.0 ml host cells ($OD_{600}=1.0$) to two 50 ml conical tube.

Inoculate with 2.5×10^5 pfu per conical tube.

5 Incubate 37°C. 20 minutes.

Add top agar to each tube to a final volume of 45 ml.

Plate the tube across five 150 mm plates.

Incubate 37°C. 6-8 hours or until plaques are about pin-head in size.

10 Overlay with 8-10 ml SM Buffer and place at 4°C overnight (with gentle rocking if possible).

Harvest Phage

Recover phage suspension by pouring the SM buffer off each plate into a 50 ml conical tube.

15 Add 3 ml chloroform, shake vigorously and incubate at room temperature, 15 minutes.

Centrifuge at 2K rpm, 10 minutes to remove cell debris.

Pour supernatant into a sterile flask, add 500 μ l chloroform.

Store at 4°C.

20 **Titer Amplified Library**

Make serial dilutions:

10^{-5} = 1 μ l amplified phage in 1 ml SM Buffer

10^{-6} = 1 μ l of the 10^{-5} dilution in 1 ml SM Buffer

Add 200 μ l host (in 10 mM $MgSO_4$) to two tubes.

25 Inoculate one with 10 μ l 10^{-5} dilution (10^{-5}).

Inoculate the other with 1 μ l 10^{-6} dilution (10^{-6}).

Incubate 37°C, 15 minutes.

Add about 3 ml 48°C top agar.

[50ml stock containing 150 μ l IPTG (0.5M) and 375 μ l X-GAL (350 mg/ml)]

30 Plate on 100 mm plates and incubate 37°C, overnight.

Excise the ZAP II library to create the pBluescript library according to manufacturers protocols (Stratagene).

5

Example 2

Activated Calf Thymus DNA Polymerase Assay

Streak out the clone to isolation:

1. Inoculate 5ml LB/Amp/Meth/Kan culture with isolated clone
2. Grow to turbidity
- 10 3. Inoculate a 50ml culture of LB/Amp/Meth/Kan
Grow to OD600 of 0.7 to 0.9; induce culture with IPTG at a final concentration of 1mM for 3 hours
Centrifuge at 4500 RPM for 20 minutes and discard supernate
Resuspend pellet in 3mls of 20mM Tris pH 8.0 and sonicate twice for 1 minute each
15 Microcentrifuge 1ml of sonicate for 30 minutes at 4°C
Remove 1ul of the sonicate supernatent and add to 10µl of the following Activated Calf Thymus Reaction Cocktail in a 0.5ml eppendorf:
5 units/ml activated calf thymus DNA (Pharmacia 27-4575-01)
1mMDTT
20 40mg/ml BSA
50uM dATP, 50uM dCTP, 50uM dGTP, 5uM dTTP
50mM Tris pH 7.6
5mM MgCl₂
50µCi/ml H³-dTTP
25 bring to volume with H₂O
Incubate at 70°C for 10-30 minutes
Stop reaction by cooling the tube
Spot sample onto Whatman DE-81 filter paper (catalog#3658-323)
Dry completely
30 Wash filters in 2X SSC five times for 2 minutes each
Final wash in 100% ethanol to remove most of remaining water

Allow the filters to dry to completion

Count incorporation of H^3 -dTTP using a scintillation counter

- 5 The incorporation of nucleotides by the polymerase is proportional to counts, by at least five fold over background. (Maki.H, et al. J.Biol.Chem. (1988) 263:6570-6578 and Tabor, et al. US Patent 4795699).

Example 3

10 PCR Screening

Polymerase sequences from *Thermococcus litoralis*, *Pyrococcus* GB-D (Deep Vent), and *Pyrococcus furiosus* were scanned to determine conserved regions. The following nucleic acid sequences were identified and corresponding amino acid sequences were utilized to derive degenerate oligonucleotide primers to be used in downstream screening:

- 15 *Thermococcus litoralis*: 37-45, 1045-1051
Pyrococcus GB-D (Deep Vent): 37-45, 1042-1049
Pyrococcus furiosus: 37-45, 505-512

- 20 The following corresponding amino acid sequences were used to produce degenerate oligonucleotide primers:

YIYALL^K/RDD
WY^C/SKECAE

- 25 The primers have been labeled Poldgen1 forward and Poldgen2 reverse:

Poldgen1 forward (26mer): 5'-TA^C/TAT^A/TTA^C/TGCTCT^C/TCTCA^A/GAGATGA-3'
Poldgen2 reverse (23mer): 5'-TC^A/TGC^A/GCA^C/TTC^C/TTTACA^A/GTACCA-3'

- 30 These primers were used to amplify potential polymerase genes directly from genomic DNA (Template DNA).

100 μ l PCR conditions:

1 μ l Poldgen1 forward (500ng/ μ l)

1 μ l Poldgen2 reverse (500ng/ μ l)

5 1 μ l 25mM dNTP mix

1 μ l Template DNA (~100ng/ μ l)

1 μ l TaqPlus Polymerase (Stratagene)

10 μ l 10X low salt reaction buffer (Stratagene)

85 μ l H₂O

10

Number of Cycles	Temperature	Time
2	95°C 42°C 72°C	30 seconds 30 seconds 2 minutes, 30 seconds
30	95°C 50°C 72°C	30 seconds 30 seconds 2 minutes, 30 seconds
1	72°C	10 minutes

15

PCR products (1.4kb bands from both organisms) were phenol chloroform extracted (ref. Maniatis) and cloned using the TA cloning system into the pGemT PCR Cloning Vector (Promega) using the following ligation reaction:

20 0.5 μ l pGemT Cloning Vector (50ng/ μ l)

2 μ l PCR Product (~1000ng/ μ l)

2 μ l rATP (10mM)

2 μ l 10X T4 Ligase Buffer

1 μ l T4 Ligase

25 12.5 μ l H₂O

Incubate 4°C overnight.

2.5µl of the above reaction was transformed into XL1-Blue MRF⁺ competent cells (Stratagene)

1.4kb PCR products were also restriction analyzed using the appropriate restriction enzymes:

Potential clones were verified by restriction analysis and sequenced.

BLASTX and BLASTN database comparisons of the sequences indicated whether the sequences were homologous to the nucleic acid sequence of a known polymerase from another organism. Amplification primers were then generated to both ends of the known polymerase gene, and were used in an amplification reaction on the genomic DNA in an attempt to pull out a full length polymerase gene from this organism. These primers include restriction sites and a new Ribosome Binding Site for downstream processing of the gene:

PCR Conditions:

15	1µl	forward primer (250ng/µl)
	1µl	reverse primer (250ng/µl)
	1µl	25mM dNTP
	1µl	template DNA (100ng/µl)
	1µl	Taq polymerase
20	10µl	10X Taq Buffer
	85µl	H ₂ O

Number of Cycles	Temperature	Time
2	95°C 42°C 72°C	30 seconds 30 seconds 2 minutes, 30 seconds
30	95°C 50°C 72°C	30 seconds 30 seconds 2 minutes, 30 seconds

GENE LIBRARY SCREENING

PCR products generated in the above reactions (using degenerate primers) were used to make long "run off" single stranded DNA probes using (P^{32} as a label).

The genomic library was screened using the single stranded (P^{32} labeled probe. Hybridization conditions for these screenings were as per Maniatis (maximum stringency for aqueous solutions; 68°C in rolling hybridization chamber).

All positive clones were then excised into pBluescript SK and sequenced.

Example 4

Expression

Positive clones were identified and isolated from the genomic library by the above methods. DNA from the clones were then used as templates in a 100 ul PCR reaction. DNA encoding the enzymes of the present invention were initially amplified from a pBluescript vector containing the DNA by the PCR technique. The amplified sequences were then inserted into a PQE vector and the enzyme was expressed according to the protocols set forth herein.

The pQE vector (Qiagen, Inc. Chatsworth, CA) encodes antibiotic resistance (Amp^r), a bacterial origin of replication (ori), an IPTG-regulatable promoter operator (P/O), a ribosome binding site (RBS), a 6-His tag and restriction enzyme sites.

The pQE vector was digested with the appropriate restriction enzymes. The amplified sequences were ligated into the respective pQE vector and inserted in frame with the sequence

encoding for the RBS. The ligation mixture was then used to transform the E. coli strain M15/pREP4 (Qiagen, Inc.) by electroporation. M15/pREP4 contains multiple copies of the plasmid pREP4, which expresses the lacI repressor and also confers kanamycin resistance (Kan^r).
5 Transformants were identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies were selected. Plasmid DNA was isolated and confirmed by restriction analysis. Clones containing the desired constructs were grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 µg/ml) and Kan (25 µg/ml). The O/N culture was used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells were grown to an
10 optical density 600 (O.D.⁶⁰⁰) of between 0.4 and 0.6. IPTG ("Isopropyl-B-D-thiogalactopyranoside") was then added to a final concentration of 1 mM. IPTG induces by inactivating the lacI repressor, clearing the P/O leading to increased gene expression. Cells were grown an extra 3 to 4 hours. Cells were then harvested by centrifugation.

Numerous modifications and variations of the present invention are possible in light of
15 the above teachings and, therefore, within the scope of the appended claims, the invention may be practiced otherwise than as particularly described. It is to be understood that, while the invention has been described with reference to the above detailed description, the foregoing description is intended to illustrate, but not to limit, the scope of the invention. Other aspects, advantages, and modifications of the invention are within the scope of the following claims. All
20 publications, patent applications, patents, and other referenced mentioned herein are incorporated by reference in their entirety.

Claims

What is claimed is:

- 1 1. Substantially pure polypeptide having an amino acid sequence selected from the group
2 consisting of SEQ ID NO:2, 4, 6, 8, 10, and 12.
- 1 2. An isolated polynucleotide sequence encoding a polypeptide of claim 1.
- 1 3. An isolated polynucleotide selected from the group consisting of:
2 a) SEQ ID NO:1, 3, 5, 7, 9, 11;
3 b) SEQ ID NO:1, 3, 5, 7, 9, 11 wherein T can also be U;
4 c) nucleic acid sequences complementary to a) and b); and
5 d) fragments of a), b), or c) that are at least 15 bases in length and that will hybridize
6 to DNA which encodes the amino acid sequences of SEQ ID NO:1, 3, 5, 7, 9, or
7 11 under moderate to highly stringent conditions.
- 1 4. The polynucleotide of claim 2, wherein the polynucleotide is isolated from a prokaryote.
- 1 5. An expression vector including the polynucleotide of claim 2.
- 1 6. The vector of claim 5, wherein the vector is a plasmid.
- 1 7. The vector of claim 5, wherein the vector is a virus-derived.
- 1 8. A host cell stably transformed with the vector of claim 5.
- 1 9. The host cell of claim 8, wherein the cell is prokaryotic.
- 1 10. The host cell of claim 8, wherein the cell is eukaryotic.
- 1 11. Antibodies that bind to a polypeptide of claim 1 or to fragments thereof.

- 1 12. The antibodies of claim 11, wherein the antibodies are polyclonal.
- 1 13. The antibodies of claim 11, wherein the antibodies are monoclonal.
- 1 14. An enzyme selected from the group consisting of:
2 a) an enzyme comprising an amino acid sequence which is at least 70% identical to
3 the amino acid sequence set forth in SEQ ID NO:2, 4, 6, 8, 10, and 12; and
4 b) an enzyme which comprises at least 30 consecutive amino acid residues
homologous with an enzyme of a).
- 1 15. A method for producing a polypeptide comprising:
2 a) culturing the host cells of claim 8;
3 b) expressing from the host cell of claim 8 a polypeptide encoded by said DNA; and
4 c) isolating the polypeptide.
- 1 16. A process for producing a cell comprising transforming or transfecting the cell with the
2 vector of claim 5 such that the cell expresses the polypeptide encoded by the DNA
3 contained in the vector.
- 1 17. An isolated polynucleotide having at least 70% identity to a member selected from the
2 group consisting of:
3 a) a polynucleotide encoding an enzyme comprising an amino acid sequence selected
4 from the group of amino acid sequences set forth in SEQ ID NOS:2, 4, 6, 8, 10,
5 and 12;
6 b) a polynucleotide which is complementary to the polynucleotide of a); and
7 c) a polynucleotide comprising at least 15 bases of the polynucleotide of a) or b).

- 1 18. An oligonucleotide probe that hybridizes to a nucleic acid target region corresponding to
2 a region selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9 and 11.
- 1 19. The probe of claim 18, wherein the probe is from 10 to about 150 nucleotides in length.
- 1 20. The probe of claim 18, wherein the probe is from 10 to about 100 nucleotides in length.
- 1 21. The probe of claim 18, wherein the probe is from 10 to about 50 nucleotides in length.
- 1 22. The probe of claim 18, wherein the probe is from 10 to about 30 nucleotides in length.
- 1 23. The probe of claim 18, wherein the probe is from 10 to about 15 nucleotides in length.
- 1 24. The oligonucleotide probe of claim 18, wherein the probe comprises a segment of 10
2 contiguous bases which is at least 70% complementary to a target sequence of 10
3 contiguous nucleotides present in the target region.
- 1 25. The probe of claim 18, wherein the probe is detectably labeled.
- 1 26. The probe of claim 25, wherein the detectable label is selected from the group consisting
2 of a radioisotope, a fluorescent compound, a bioluminescent compound, a
3 chemiluminescent compound, a metal chelator or an enzyme.
- 1 27. An antibody that binds to a polypeptide as set forth in SEQ ID NOs:2, 4, 6, 8, 10 and 12.
- 1 28. The antibody of claim 27, wherein the antibody is polyclonal.
- 1 29. The antibody of claim 27, wherein the antibody is monoclonal.
- 1 30. The antibody of claim 27, wherein the antibody is detectably labeled.

1/28

FIGURE 1

Ammonifex degensii DNA Polymerase (3py1)

1

(SEQ ID NO:1) GTG AAG GGA AAA ACC TTG CTC CTT TTG GAC GGC TCG AGC ATA GCC TAC
CGG

(SEQ ID NO:2) Val Lys Gly Lys Thr Leu Leu Leu Leu Asp Gly Ser Ser Ile Ala Tyr
Arg

GCC TTT TTC GCC CTT CCC TCC CTC CGC ACC CGT ACC GGC CTG CCC ACC GGT
Ala Phe Phe Ala Leu Pro Ser Leu Arg Thr Arg Thr Gly Leu Pro Thr Gly

GCC GTG TAC GGC TTT ACC TCC ATG CTC TTC AAA GTG CTG GAA GAA AGG CGT
Ala Val Tyr Gly Phe Thr Ser Met Leu Phe Lys Val Leu Glu Glu Arg Arg

CCC ACG GCC ATA GTG GCG GCT TTC GAT AAA AGC AAG ACC ACC TTC CGG CAC
Pro Thr Ala Ile Val Ala Ala Phe Asp Lys Ser Lys Thr Thr Phe Arg His

GCC CTG GCG GAG ACC TAC AAG GCC CAC CGC CCC GCC ACT CCG GAT GAA CTG
Ala Leu Ala Glu Thr Tyr Lys Ala His Arg Pro Ala Thr Pro Asp Glu Leu

CGC CAG CAG TTC AAC CTC ATC AAG GAA GTG CTG ACT GCC CTC AAC GTT CCG
Arg Gln Gln Phe Asn Leu Ile Lys Glu Val Leu Thr Ala Leu Asn Val Pro

GTA GTG GAA AAG GAG GGT TTT GAG GCC GAC GAC CTC ATC GGC ACT CTG GTA
Val Val Glu Lys Glu Gly Phe Glu Ala Asp Asp Leu Ile Gly Thr Leu Val

GAC CGG GCG GAA AAA GAG GGT TGG CAG TGC CTT ATC GTC ACC GGC GAC CTC
Asp Arg Ala Glu Lys Glu Gly Trp Gln Cys Leu Ile Val Thr Gly Asp Leu

GAC GCC CTG CAG CTG GTT TCC CCC CTC ACC ACC GTC GTC CTC ATG CGC AAG
Asp Ala Leu Gln Leu Val Ser Pro Leu Thr Thr Val Val Leu Met Arg Lys

GGG ATA AGC GAA ATA GCG GTC TTT AAC GAG GCG GAG GTG AAA CGC CGC TTC
Gly Ile Ser Glu Ile Ala Val Phe Asn Glu Ala Glu Val Lys Arg Arg Phe

GGC GTC ACA CCC CGC CAA CTC CCC GAC TTC AAA GCC TTG GCC GGA GAT GCC
Gly Val Thr Pro Arg Gln Leu Pro Asp Phe Lys Ala Leu Ala Gly Asp Ala

TCG GAC AAC ATC CCC GGG CTT CCG GGC ATA GGG CCC AAA ACT GCC TCC CGT
Ser Asp Asn Ile Pro Gly Leu Pro Gly Ile Gly Pro Lys Thr Ala Ser Arg

CTG CTA CAG TCC CAC CAG AGC CTG GAG AAA TTG CTG GAG AGC AAG GAA TTT

2/28

Fig. 1 CONTINUED

Leu Leu Gln Ser His Gln Ser Leu Glu Lys Leu Leu Glu Ser Lys Glu Phe

TTT CCG GGC AAG CTG CGC GAA ACC TTA GAA AGG CAC AAG GAA GAA GCG GTT
Phe Pro Ala Lys Leu Arg Glu Thr Leu Glu Arg His Lys Glu Glu Ala Val

TTG GCC AAG AAA CTG GCC CTC ATC CGC CGC GAT GTG CCG CTG GAA GAG GAG
Leu Ala Lys Lys Leu Ala Leu Ile Arg Arg Asp Val Pro Leu Glu Glu Glu

ATC ATC CGG CCC TGG CCG GGA CCC AAC ATT TTA GCC ACG CTG GAG GTC TTC
Ile Ile Arg Pro Trp Pro Gly Pro Asn Ile Leu Ala Thr Leu Glu Val Phe

TCG CGC CTG GAA TTC CGC ACC TTG GCC AAG AGA TTC CTC GAG CTT TTC CCC
Ser Arg Leu Glu Phe Arg Thr Leu Ala Lys Arg Phe Leu Glu Leu Phe Pro

GAG GCA CGC CTC CTG TCC GCC AGT GGC CTT ACC CCC TCC GCT GTC CGC GTA
Glu Ala Arg Leu Leu Ser Ala Ser Gly Leu Thr Pro Ser Ala Val Arg Val

AAG GTA GAA AGA CCC GAA GAA CTA GAA AGA CTG GGG GAA GAG CTC GGA AGG
Lys Val Glu Arg Pro Glu Glu Leu Glu Arg Leu Gly Glu Glu Leu Gly Arg

CAA GAA TTT GCG GCC CTG GCT TAC CCC CCC GTT CTT CGG CGC AAA GCC ACT
Gln Glu Phe Ala Ala Leu Ala Tyr Pro Pro Val Leu Arg Arg Lys Ala Thr

TCT TCT TTC TTG GCT CTC TGT CTG GGA GGG GAA AAG GTC TTC CTG CTG GAA
Ser Ser Phe Leu Ala Leu Cys Leu Gly Gly Glu Lys Val Phe Leu Leu Glu

GGG CCG GAG GTG CTC AAG AGC TTC TTC CGG CTG CTC GAA GAA AAG GGA GGT
Gly Pro Glu Val Leu Lys Ser Phe Phe Arg Leu Leu Glu Glu Lys Gly Gly

CTT GTC AGT ACC TAC GAC GCT AAA TCC TGC CTT CAC GCC CTG GAA CCT TAC
Leu Val Ser Thr Tyr Asp Ala Lys Ser Cys Leu His Ala Leu Glu Pro Tyr

GGC TTC AAG CCC GAA ATG ATC GGG TTT GAC GTC CTG CTG GCA GCC TAC CTG
Gly Phe Lys Pro Glu Met Ile Gly Phe Asp Val Leu Leu Ala Ala Tyr Leu

GTG AAC CCC GCC GCC AAC AAC GAA CTG GGG GCG ATC GCC TTC GAG CAC GCG
Val Asn Pro Ala Ala Asn Asn Glu Leu Gly Ala Ile Ala Phe Glu His Ala

GGC TTT ATG CTC TCC CCG GGA GCA GAG CTC CCG GAA AAA GCC CAG GCG ATC
Gly Phe Met Leu Ser Pro Gly Ala Glu Leu Pro Glu Lys Ala Gln Ala Ile

TAC CAG CTC ACC CCC ATC CTA AAA AGT AAG ATT AAG CTT CAG GAA CAG GAG
Tyr Gln Leu Thr Pro Ile Leu Lys Ser Lys Ile Lys Leu Gln Glu Gln Glu

3/28

Fig. 1 CONTINUED

TAC CTT TAT TAC TCC GTG GAG CTT CCC TTA GCC GCC GTC TTG GCC GAC ATG
Tyr Leu Tyr Tyr Ser Val Glu Leu Pro Leu Ala Ala Val Leu Ala Asp Met

GAG AAA GTC GGG GTG AAA GTT TCG GAG GAA AGG CTG CGT TCT CTC TCC AAG
Glu Lys Val Gly Val Lys Val Ser Glu Glu Arg Leu Arg Ser Leu Ser Lys

GAG CTG GGA GAG CAG CTG GCT CAG CTT TCC GAG GAA ATC TAT AAG CTC GCC
Glu Leu Gly Glu Gln Leu Ala Gln Leu Ser Glu Glu Ile Tyr Lys Leu Ala

GGC GAG CGC TTC AAC CTG AAT TCC CCC CGC CAG CTC GGC TAC ATC CTG TTC
Gly Glu Arg Phe Asn Leu Asn Ser Pro Arg Gln Leu Gly Tyr Ile Leu Phe

GAG AAG TTG GGA CTC AAA CCG GTC AAG AAG ACC AAA ACC GGC TAC TCC ACC
Glu Lys Leu Gly Leu Lys Pro Val Lys Lys Thr Lys Thr Gly Tyr Ser Thr

GAC GCT TCG GTC CTA GAA AAG CTA GCC GAG CAC GAG ATC GTG GCT AAG GTG
Asp Ala Ser Val Leu Glu Lys Leu Ala Glu His Glu Ile Val Ala Lys Val

CTC GTC TAC CGG CAG CTG GCC AAA CTA AAG AGC ACT TAC ACC GAC GCA CTT
Leu Val Tyr Arg Gln Leu Ala Lys Leu Lys Ser Thr Tyr Thr Asp Ala Leu

CCA GAG CTC ATC GAC CCG GCC ACC GGG CGC CTG CAC ACC ACC TTC TTG CAG
Pro Glu Leu Ile Asp Pro Ala Thr Gly Arg Leu His Thr Thr Phe Leu Gln

GCA GGG ACG GCA ACG GGA AGA CTG GCC TCC GCC GAG CCC AAC CTG CAG AAC
Ala Gly Thr Ala Thr Gly Arg Leu Ala Ser Ala Glu Pro Asn Leu Gln Asn

ATT CCC GTA CGC GAT TCT CTG GGA AGG CGC ATC CGG CAG GCC TTC GTG GCT
Ile Pro Val Arg Asp Ser Leu Gly Arg Arg Ile Arg Gln Ala Phe Val Ala

GAG GGC CCC GAC TAC GTG CTA CTA AGC GCC GAC TAC TCC CAG ATA GAG CTT
Glu Gly Pro Asp Tyr Val Leu Leu Ser Ala Asp Tyr Ser Gln Ile Glu Leu

CGG GTC TTG GCC CAC CTT TCC GAA GAT CCG GGG CTG TGT GAG GCC TTT GTT
Arg Val Leu Ala His Leu Ser Glu Asp Pro Gly Leu Cys Glu Ala Phe Val

AAA GGA GAA GAC ATT CAC GCC CGC ACG GCG GCC GAG ATC TTC GGC GTT TCT
Lys Gly Glu Asp Ile His Ala Arg Thr Ala Ala Glu Ile Phe Gly Val Ser

CCT CAG GAA GTG ACG CCG GAG ATG CGG GCC AAG GCC AAG GTG GTA AAC TTC
Pro Gln Glu Val Thr Pro Glu Met Arg Ala Lys Ala Lys Val Val Asn Phe

GGG ATC GTT TAC GGC ATG AGC GAT TAC GGC CTT TCC CAG GAG CTC AAG ATC

4/28

Fig. 1 Continued

Gly Ile Val Tyr Gly Met Ser Asp Tyr Gly Leu Ser Gln Glu Leu Lys Ile
GAG CCC GGC GAG GCG CAC GAG TAT ATA GAA CGG TAC TTC CGG CGC TAT CCG
Glu Pro Gly Glu Ala His Glu Tyr Ile Glu Arg Tyr Phe Arg Arg Tyr Pro
CGC GTG AAG CAG TTC ATC GAG CGG GTG ATC GCC CAG GCC CGA GAG AAG GGC
Arg Val Lys Gln Phe Ile Glu Arg Val Ile Ala Gln Ala Arg Glu Lys Gly
TAC GTG ACC ACT ATT CTC AAC CGC CGC CGC TAC ATC CCT GAA ATA CTG AGC
Tyr Val Thr Thr Ile Leu Asn Arg Arg Arg Tyr Ile Pro Glu Ile Leu Ser
AGC AAC CGC AAC CAG CGT CAG CTG GGG GAG CGC CTG GCC ATC AAC ACC ACC
Ser Asn Arg Asn Gln Arg Gln Leu Gly Glu Arg Leu Ala Ile Asn Thr Thr
ATT CAA GGA AGT GCG GCC GAT CTT ATA AAA AAG GCC ATG GTG GAC ATC CAC
Ile Gln Gly Ser Ala Ala Asp Leu Ile Lys Lys Ala Met Val Asp Ile His
CGG CAA CTG AAA GGG CAA GGA TTT AAA TGC CGG ATG ATC CTC CAG GTG CAC
Arg Gln Leu Lys Gly Gln Gly Phe Lys Cys Arg Met Ile Leu Gln Val His
GAC GAA CTC CTC TTC GAG GTG CCT AAA GAA GAA CTG GAA AAG GTG GCA CCT
Asp Glu Leu Leu Phe Glu Val Pro Lys Glu Glu Leu Glu Lys Val Ala Pro
ATA ATA AAA AGC ACC ATG GAG CAA GCC TTA CCT TTT AAG GTT CCC ATA AAG
Ile Ile Lys Ser Thr Met Glu Gln Ala Leu Pro Phe Lys Val Pro Ile Lys
GCC AAC CTC AAG GTA GGG CCT AAC TGG CAA GAC ATG GAA GAG TAC GAG GTG
Ala Asn Leu Lys Val Gly Pro Asn Trp Gln Asp Met Glu Glu Tyr Glu Val
2602
GAA TGA Glu End

5/28

Fig 2.

FIGURE 2

Pyrolobus fumarius DNA Polymerase (IPY2)

(SEQ ID NO:3)ATG ACG GAA GTT GTA TTC ACG GTT TTA GAC TCT AGC TAC GAG GTT GTT GGT

(SEQ ID NO:4)Met Thr Glu Val Val Phe Thr Val Leu Asp Ser Ser Tyr Glu Val Val Gly

AAA GAG CCT CAG GTA ATC ATA TGG GGT ATT GCT GAG AAC GGC GAG AGG GTA
Lys Glu Pro Gln Val Ile Ile Trp Gly Ile Ala Glu Asn Gly Glu Arg Val

GTC CTC ATT GAC AGG TCT TTT CGC CCA TAC TTC TAT GCG CTG CTT GCA CCG
Val Leu Ile Asp Arg Ser Phe Arg Pro Tyr Phe Tyr Ala Leu Leu Ala Pro

GGC GCC GAT CCT AAG CAG GTA GCA CAA CGT ATT CGT GCA TTG AGT AGG CCA
Gly Ala Asp Pro Lys Gln Val Ala Gln Arg Ile Arg Ala Leu Ser Arg Pro

AAG AGC CCG ATT ATA GGT GTA GAG GAT GAC AAG AGG AAG TAC TTC GGG AGG
Lys Ser Pro Ile Ile Gly Val Glu Asp Asp Lys Arg Lys Tyr Phe Gly Arg

CCT CGT AGG GTC TTA CGT ATT CGC ACC GTG CTA CCC GAG GCT GTT AGG GAG
Pro Arg Arg Val Leu Arg Ile Arg Thr Val Leu Pro Glu Ala Val Arg Glu

TAT CGC GAA CTC GTA AAG AAC GTT GAT GGT GTT GAG GAT GTT CTA GAG GCG
Tyr Arg Glu Leu Val Lys Asn Val Asp Gly Val Glu Asp Val Leu Glu Ala

GAT ATA CGC TTC GCT ATG CGC TAT CTC ATA GAT CAC GAT CTA TTT CCT TTC
Asp Ile Arg Phe Ala Met Arg Tyr Leu Ile Asp His Asp Leu Phe Pro Phe

ACC TGG TAC CGT GTA GAG GCT GAG CCC CTC GAG AAC AAG ATG GGC TTC CGT
Thr Trp Tyr Arg Val Glu Ala Glu Pro Leu Glu Asn Lys Met Gly Phe Arg

GTC GAC AAG GTA TAC CTG GTT AAG AGC AGG CCG GAG CCA CTT TAT GGT GAG
Val Asp Lys Val Tyr Leu Val Lys Ser Arg Pro Glu Pro Leu Tyr Gly Glu

6/28

Fig. 2 CONTINUED

- -

GCT CTC GCA CCA ACC AAG CTT CCC GAT CTT AGG ATA CTC GCG TTC GAT ATT
Ala Leu Ala Pro Thr Lys Leu Pro Asp Leu Arg Ile Leu Ala Phe Asp Ile

GAA GTT TAT AGC AAG CAA GGG TCG CCG CGT CCA GAG CGC GAT CCT GTA ATA
Glu Val Tyr Ser Lys Gln Gly Ser Pro Arg Pro Glu Arg Asp Pro Val Ile

GTG ATA GCT GTG AAG ACT GAC GAT GGC GAT GAG GTG CTA TTC ATT GCA GAG
Val Ile Ala Val Lys Thr Asp Asp Gly Asp Glu Val Leu Phe Ile Ala Glu

GGC AAA GAC GAT CGA AAA CCG ATA CGC GAG TTT GTA GAG TAC GTG AAG AGG
Gly Lys Asp Asp Arg Lys Pro Ile Arg Glu Phe Val Glu Tyr Val Lys Arg

TAT GAC CCC GAC ATA ATA GTC GGT TAT AAC AAC AAT CAT TTC GAT TGG CCT
Tyr Asp Pro Asp Ile Ile Val Gly Tyr Asn Asn Asn His Phe Asp Trp Pro

TAT CTT TTG AGG CGC GCC CGC ATC CTA GGC ATA AAG CTT GAT GTG ACT AGA
Tyr Leu Leu Arg Arg Ala Arg Ile Leu Gly Ile Lys Leu Asp Val Thr Arg

AGA GTT GGC GCC GAG CCC ACC ACT AGC GTA CAT GGG CAC GTC TCT GTC CCT
Arg Val Gly Ala Glu Pro Thr Thr Ser Val His Gly His Val Ser Val Pro

GGC AGG CTT AAC GTA GAT CTG TAC GAC TAT GCC GAA GAG ATG CCA GAG ATC
Gly Arg Leu Asn Val Asp Leu Tyr Asp Tyr Ala Glu Glu Met Pro Glu Ile

AAG ATA AAG AGT CTC GAG GAG GTC GCA GAG TAT CTA GGC GTG ATG AAG AAG
Lys Ile Lys Ser Leu Glu Glu Val Ala Glu Tyr Leu Gly Val Met Lys Lys

AGT GAA CGC CTT ATC ATC AAT TGG TGG GAG ATT CCA GAC TAT TGG GAC GAC
Ser Glu Arg Val Ile Ile Asn Trp Trp Glu Ile Pro Asp Tyr Trp Asp Asp

CCG AAG AAG AGA CCA CTA TTA CTG CAA TAC GCG CGC GAC GAT GTC CGC GCT
Pro Lys Lys Arg Pro Leu Leu Leu Gln Tyr Ala Arg Asp Asp Val Arg Ala

ACT TAC GGC TTA GCC GAG AAG ATA TTG CCG TTT GCT ATC CAG TTG TCG TAC

7/28

Fig. 2 CONTINUED

- -

Thr Tyr Gly Leu Ala Glu Lys Ile Leu Pro Phe Ala Ile Gln Leu Ser Tyr

GTA ACA GGT CTC CCA CTA GAC CAG GTA GGT GCG ATG AGT GTT GGC TTT CGA
Val Thr Gly Leu Pro Leu Asp Gln Val Gly Ala Met Ser Val Gly Phe Arg

CTT GAA TGG TAC CTG ATA CGC GCG GCG TTT AAG ATG AAA GAG CTT GTG CCG
Leu Glu Trp Tyr Leu Ile Arg Ala Ala Phe Lys Met Lys Glu Leu Val Pro

AAC CGC GTT GAG CGC CCA GAA GAG ACT TAC CGT GGC GCT ATA GTT CTT GAG
Asn Arg Val Glu Arg Pro Glu Glu Thr Tyr Arg Gly Ala Ile Val Leu Glu

CCG TTG AGA GGC GTG CAC GAG AAT ATA GCC GTA CTC GAC TTT AGC TCG ATG
Pro Leu Arg Gly Val His Glu Asn Ile Ala Val Leu Asp Phe Ser Ser Met

TAC CCA AAC ATC ATG ATA AAG TAC AAT GTT GGT CCT GAC ACG CTT GTG AGG
Tyr Pro Asn Ile Met Ile Lys Tyr Asn Val Gly Pro Asp Thr Leu Val Arg

CCT GGT GAA AAG TGT GGC GAG TGT GGT TGC TGG GAG GCC CCG GAG GTC AAG
Pro Gly Glu Lys Cys Gly Glu Cys Gly Cys Trp Glu Ala Pro Glu Val Lys

CAC AGG TTC CGT AGG TGT CCG CCC GGC TTC TTC AAG ACA GTT CTT GAG AGG
His Arg Phe Arg Arg Cys Pro Pro Gly Phe Phe Lys Thr Val Leu Glu Arg

CTG TTA GAG CTT CGT AAG CGT GTG CGT GCT GAA ATG AAG AAG TAT CCT CCG
Leu Leu Glu Leu Arg Lys Arg Val Arg Ala Glu Met Lys Lys Tyr Pro Pro

GAT AGC CCA GAA TAT CGA CTG TTG GAT GAA AGG CAG AAG GCG TTG AAG GTT
Asp Ser Pro Glu Tyr Arg Leu Leu Asp Glu Arg Gln Lys Ala Leu Lys Val

CTT GCA AAC GCT AGT TAC GGC TAC ATG GGT TGG AGC GGC GCT AGG TGG TAT
Leu Ala Asn Ala Ser Tyr Gly Tyr Met Gly Trp Ser Gly Ala Arg Trp Tyr

TGC AGG GAG TGC GCA AAG GCT GTC ACG GCT TGG GGT AGG CAC CTC ATA CGC
Cys Arg Glu Cys Ala Lys Ala Val Thr Ala Trp Gly Arg His Leu Ile Arg

8/28

Fig. 2 CONTINUED

--

ACC GCC ATC AAC ATA GCT CGT AAA CTA GGC CTC AAG GTG ATC TAC GGT GAC
Thr Ala Ile Asn Ile Ala Arg Lys Leu Gly Leu Lys Val Ile Tyr Gly Asp

ACA GAT TCG CTC TTC GTG ACC TAT GAT CCG GAG AAG GTG GAA AAT TTC ATC
Thr Asp Ser Leu Phe Val Thr Tyr Asp Pro Glu Lys Val Glu Asn Phe Ile

AAA ATT ATA AAG GAG GAG CTG GGG TTC GAA ATC AAG CTA GAG AAG GTG TAC
Lys Ile Ile Lys Glu Glu Leu Gly Phe Glu Ile Lys Leu Glu Lys Val Tyr

AAA CGC TTA TTC TTT ACA GAG GCT AAG AAG AGG TAC GCT GGC CTT CTC GAG
Lys Arg Leu Phe Phe Thr Glu Ala Lys Lys Arg Tyr Ala Gly Leu Leu Glu

GAC GGA CGT ATA GAT ATT GTC GGT TTC GAG GCT GTA CGT GGC GAT TGG TGT
Asp Gly Arg Ile Asp Ile Val Gly Phe Glu Ala Val Arg Gly Asp Trp Cys

GAA CTC GCC AAG GAG GTT CAG ACT AAG GTT GTC GAA ATA GTA TTG AAG ACG
Glu Leu Ala Lys Glu Val Gln Thr Lys Val Val Glu Ile Val Leu Lys Thr

AGT GAG GTG AAC AAG GCT GTA GAG TAC GTC AGG AAG ATT GTG AAA GAG TTG
Ser Glu Val Asn Lys Ala Val Glu Tyr Val Arg Lys Ile Val Lys Glu Leu

GAG GAG GGC AAG GTT CCC ATA GAG AAG CTT GTA ATC TGG AAG ACC CTT AGT
Glu Glu Gly Lys Val Pro Ile Glu Lys Leu Val Ile Trp Lys Thr Leu Ser

AAG CGT CTT GAG GAG TAC ACA ACG GAG GCA CCA CAC GTC GTT GCA GCG AAG
Lys Arg Leu Glu Glu Tyr Thr Thr Glu Ala Pro His Val Val Ala Ala Lys

AGG ATG CTG TCA GCA GGC TAC CGG GTA AGC CCA GGC GAC AAG ATA GGG TAT
Arg Met Leu Ser Ala Gly Tyr Arg Val Ser Pro Gly Asp Lys Ile Gly Tyr

GTA ATA GTG AAG GGT GGT GGC CGT ATC AGT CAA AGA GCA TGG CCA TAC TTC
Val Ile Val Lys Gly Gly Gly Arg Ile Ser Gln Arg Ala Trp Pro Tyr Phe

9/28

Fig. 2. CONTINUED

ATG GTC AAG GAT CCT AGC CAG ATA GAC GTG ACC TAC TAT GTT GAC CAC CAA
Met Val Lys Asp Pro Ser Gln Ile Asp Val Thr Tyr Tyr Val Asp His Gln

ATC ATC CCG GCT GCA TTG AGA ATA CTG GGC TAC TTT GGC ATC ACC GAG AAG
Ile Ile Pro Ala Ala Leu Arg Ile Leu Gly Tyr Phe Gly Ile Thr Glu Lys

AAG CTG AAA GCA AGT GCA ACT GGG CAG AAG ACT CTC TTC GAC TTT CTA GCC
Lys Leu Lys Ala Ser Ala Thr Gly Gln Lys Thr Leu Phe Asp Phe Leu Ala

2412

AAG AAG AGC AAG TAA

Lys Lys Ser Lys End

10/28

FIGURE 3

Archaeoglobus lithotrophicus DNA Polymerase (SPY1)

1

(SEQ ID NO:5) ATG ATA AAG GTC AAG GGC TGG CTG CTC GAT GCA GAT TAT
ATC ACC GAA AAC

(SEQ ID NO:6) Met Ile Lys Val Lys Gly Trp Leu Leu Asp Ala Asp Tyr
Ile Thr Glu Asn

GAT CGA GCC GTT ATA AGG CTA TGG TGT AAG GAT GAG GAA GGA ATA TTT ATC
Asp Arg Ala Val Ile Arg Leu Trp Cys Lys Asp Glu Glu Gly Ile Phe Ile

GCA TAC GAT CAC TCA TTC CAG CCC TAC TTT TAC GCA CTC AAA GAA GAG GGT
Ala Tyr Asp His Ser Phe Gln Pro Tyr Phe Tyr Ala Leu Lys Glu Glu Gly

ATC ACT GCC GAA GAT ATA GTG AAA ATA AAG GTT CAA ACG AAA AAA GAA GTA
Ile Thr Ala Glu Asp Ile Val Lys Ile Lys Val Gln Thr Lys Lys Glu Val

ATT ACG CCG TTA AAA GTT GAG GAA ACC ACA GCC AAA AAT CTT GGT AGG GAG
Ile Thr Pro Leu Lys Val Glu Glu Thr Thr Ala Lys Asn Leu Gly Arg Glu

GTT GAA GTT TTC AAG ATA TAT GCA AGA CAC CCT CAG CAC GTC CCC AAA CTT
Val Glu Val Phe Lys Ile Tyr Ala Arg His Pro Gln His Val Pro Lys Leu

CGT GAG GTT GTT TCG CAG TAT CTG GAG ATT AGG GAG GCA GAC ATA CCT TTT
Arg Glu Val Val Ser Gln Tyr Leu Glu Ile Arg Glu Ala Asp Ile Pro Phe

GCC TAT CGA TAC CTC ATA GAT AAA AAT CTT GCG TGT ATG GAT GGA GTT GTA
Ala Tyr Arg Tyr Leu Ile Asp Lys Asn Leu Ala Cys Met Asp Gly Val Val

ATT GAA GGC GTT GAA AGA CGT GAG AAG GGG TTG AGA TGT TAC GAA ATC AAG
Ile Glu Gly Val Glu Arg Arg Glu Lys Gly Leu Arg Cys Tyr Glu Ile Lys

AGA ATA GAA AGA GAT TCC AGA CAG GAT TTT CCC GAA CTC AAG GTT ATG GCG
Arg Ile Glu Arg Asp Ser Arg Gln Asp Phe Pro Glu Leu Lys Val Met Ala

11/28

Fig. 3 CONTINUED

TTT GAT TGC GAA ATG CTC TCA GAG GTT GGT ATG CCC GAT CCA GAG AAA GAT
Phe Asp Cys Glu Met Leu Ser Glu Val Gly Met Pro Asp Pro Glu Lys Asp

CCT ATC ATA GTC ATA TCA ATT AAA TCG GGT GAA TAC GAG GAA ATC CTC AAC
Pro Ile Ile Val Ile Ser Ile Lys Ser Gly Glu Tyr Glu Glu Ile Leu Asn

GGT GAT AAC GAG AGA GAA TTG CTT ACC AGA TTT GTC AAG ATA ATT CGC GAT
Gly Asp Asn Glu Arg Glu Leu Leu Thr Arg Phe Val Lys Ile Ile Arg Asp

ATT GAT CCC GAC ATT ATA GTT GGA TAC AAT CAG GAC AGC TTT GAC TGS CCC
Ile Asp Pro Asp Ile Ile Val Gly Tyr Asn Gln Asp Ser Phe Asp Trp Pro

TAT ATC AAG AAG AGA GCT GAG AAA CTG AGG GTT AAG CTT GAC ATC GGA AGA
Tyr Ile Lys Lys Arg Ala Glu Lys Leu Arg Val Lys Leu Asp Ile Gly Arg

GAT AGA AGC GAA CTG GCT ATC AGG GGA GGA AGA CCA AAG ATT GCT GGC AGG
Asp Arg Ser Glu Leu Ala Ile Arg Gly Gly Arg Pro Lys Ile Ala Gly Arg

TTG AAC GTG GAT CTC TAT GAT ATT GCA ATG AGG AGT CTC GAT GTA AAG GTG
Leu Asn Val Asp Leu Tyr Asp Ile Ala Met Arg Ser Leu Asp Val Lys Val

AAG AAG CTC GAA AAC GTT GCA GAG TTT CTG GGT AAG AAA ATA GAG CTT GCA
Lys Lys Leu Glu Asn Val Ala Glu Phe Leu Gly Lys Lys Ile Glu Leu Ala

GAT ATT GAA GCG AAG GAT ATC TAC AAG CAC TGG ACA TCG GGC GAC AGG GAA
Asp Ile Glu Ala Lys Asp Ile Tyr Lys His Trp Thr Ser Gly Asp Arg Glu

AGC GTA ATC AAA TAC TCC CGG CAG GAC ATC CTG CAC ACG TAC TTC ATA GCT
Ser Val Ile Lys Tyr Ser Arg Gln Asp Ile Leu His Thr Tyr Phe Ile Ala

GAA GAA TTG CTG CCA ATG CAT TAC GAA CTT TCC AGA ATG ATA CGC ATA CCT
Glu Glu Leu Leu Pro Met His Tyr Glu Leu Ser Arg Met Ile Arg Ile Pro

12/28

Fig. 3 Continued

CTC GAT GAT GTG ACA AGG AGC GGG AGA GGT AAG CAG GTT GAG TGG CTG CTG
Leu Asp Asp Val Thr Arg Ser Gly Arg Gly Lys Gln Val Glu Trp Leu Leu

TTA AGC GAA GCA CAC AAA CTT GGC GAA CTT GCA CCC AAC CCC AGA GAG ATG
Leu Ser Glu Ala His Lys Leu Gly Glu Leu Ala Pro Asn Pro Arg Glu Met

GCC GAC AGC TAT GAA GGA GCA TTC GTG CTC GAG CCC GCA AGA GGA TTG CAT
Ala Asp Ser Tyr Glu Gly Ala Phe Val Leu Glu Pro Ala Arg Gly Leu His

GAG AAC GTA ATC TGC CTG GAC TTT GCG TCC ATG TAT CCC TCA ATA ATG ATT
Glu Asn Val Ile Cys Leu Asp Phe Ala Ser Met Tyr Pro Ser Ile Met Ile

TCA TAC AAC ATC AGC CCC GAC ACG CTT GTA ATA GGC AAA TGC GAC GAT TGC
Ser Tyr Asn Ile Ser Pro Asp Thr Leu Val Ile Gly Lys Cys Asp Asp Cys

AAT GTA GCG CCG GAG GTG GGG CAC AAA TTC AGG AAA CAT CCT GAT GGT TTT
Asn Val Ala Pro Glu Val Gly His Lys Phe Arg Lys His Pro Asp Gly Phe

TTC AAA AGA ATA CTC AAA ATG CTG ATT GAG AAA AGA AGA GAA ATA AAG AAG
Phe Lys Arg Ile Leu Lys Met Leu Ile Glu Lys Arg Arg Glu Ile Lys Lys

GTT ATG AAA ACA CTT GAC TAC AAC TCG CCA GAA TAC AAG CTG CTC GAT ATA
Val Met Lys Thr Leu Asp Tyr Asn Ser Pro Glu Tyr Lys Leu Leu Asp Ile

AAG CAG GCA ACG CTG AAA GTT CTT ACA AAC TCG TTT TAC GGT TAT ACT GGG
Lys Gln Ala Thr Leu Lys Val Leu Thr Asn Ser Phe Tyr Gly Tyr Thr Gly

TGG AGT CTT GCG AGA TGG TAC TGC AAG GAG TGC GCT GAA GCT ACA ACG GCA
Trp Ser Leu Ala Arg Trp Tyr Cys Lys Glu Cys Ala Glu Ala Thr Thr Ala

TGG GGC AGA CAC TTT ATC AAA ACA TCT GCA AGA ATT GCG AAA GAG CTT GGA
Trp Gly Arg His Phe Ile Lys Thr Ser Ala Arg Ile Ala Lys Glu Leu Gly

TTT GAA GTG CTA TAT GGG GAT ACA GAT AGC ATC TTT GTT AAA AAA GAT GGA

13/28

Fig. 3 CONTINUED

Phe Glu Val Leu Tyr Gly Asp Thr Asp Ser Ile Phe Val Lys Lys Asp Gly

TTG AGC CTG GAA GAG CTC AAA AAA GAA GTT AAA AAG CTC ATA GGT AAA CTT
Leu Ser Leu Glu Glu Leu Lys Lys Glu Val Lys Lys Leu Ile Gly Lys Leu

TCG GAA GAG ATG CCA ATA CAA ATA GAG ATA GAT GAA TAC TAC GAG ACA ATA
Ser Glu Glu Met Pro Ile Gln Ile Glu Ile Asp Glu Tyr Tyr Glu Thr Ile

TTC TTC GTT GAA AAG AAA AGG TAT GCT GGA TTG ACA CAG GAT GGA AGA ATA
Phe Phe Val Glu Lys Lys Arg Tyr Ala Gly Leu Thr Gln Asp Gly Arg Ile

ATT GTA AAG GGT CTT GAA GTC AGA AGA GGC GAC TGG TGC GAG CTT GCA AAG
Ile Val Lys Gly Leu Glu Val Arg Arg Gly Asp Trp Cys Glu Leu Ala Lys

AAG ATA CAG AAA GGT GTA ATA GAA ATC ATT CTG AAG GAA AAG AAT CCT GAA
Lys Ile Gln Lys Gly Val Ile Glu Ile Ile Leu Lys Glu Lys Asn Pro Glu

AAA GCT GCT GAG TAT GTG AAA GGA GTC ATA GAG GAG ATA AAG GCA GGC AAA
Lys Ala Ala Glu Tyr Val Lys Gly Val Ile Glu Glu Ile Lys Ala Gly Lys

ATT CCG CTT GAA GAT TAT ATC ATC TAC AAG GGA TTG ACG AGA AAA CCA TCA
Ile Pro Leu Glu Asp Tyr Ile Ile Tyr Lys Gly Leu Thr Arg Lys Pro Ser

AAG TAC GAG AGT ATG CAG GCT CAC GTA AAA GCT GCC ATG AAG GCG GCA AAG
Lys Tyr Glu Ser Met Gln Ala His Val Lys Ala Ala Met Lys Ala Ala Lys

AGA GGA ATA GTA TAC ACA ATC GGC TCA AAG GTT GGT TTT GTC GTT ACA AAA
Arg Gly Ile Val Tyr Thr Ile Gly Ser Lys Val Gly Phe Val Val Thr Lys

GGT GTG GGG AAC ATA GGT GAT AGG GCT TTT CCA TCT GAT CTG ATA GAG GAC
Gly Val Gly Asn Ile Gly Asp Arg Ala Phe Pro Ser Asp Leu Ile Glu Asp

TTT GAC GGT GAA GTG ATC ACA GAT CTT GAC GGA AAC AAG TAC AAG ATC GAC
Phe Asp Gly Glu Val Ile Thr Asp Leu Asp Gly Asn Lys Tyr Lys Ile Asp

14/28

Fig. 3 CONTINUED --

AAG GAA TAC TAT ATA GAC CAT CAG GTA CTG CCA TCG GTT CTT CGA ATT CTC
Lys Glu Tyr Tyr Ile Asp His Gln Val Leu Pro Ser Val Leu Arg Ile Leu

GAG AGG TTC GGA TAC ACC GAG GCA CAG CTA AAA GGT GCT GCG GAG CAG CAA
Glu Arg Phe Gly Tyr Thr Glu Ala Gln Leu Lys Gly Ala Ala Glu Gln Gln

ACG CTA GAT GCT TTC TGG TAA
Thr Leu Asp Ala Phe Trp End

15/28

FIGURE 4

Metallosphaera prunae DNA Polymerase (23PY1)

1

(SEQ ID NO:7) ATG AGT ATA ATG GCC AGA CAG CTT ACC CTT GCT GAC TTC
TCT GGG ATC AAG

(SEQ ID NO:8) Met Ser Ile Met Ala Arg Gln Leu Thr Leu Ala Asp Phe
Ser Gly Ile Lys

AGA GAG GAA CCA GTT AAA CAG GAA GAG AAG ACG CAG GAG GAA GAG AGG CCT
Arg Glu Glu Pro Val Lys Gln Glu Glu Lys Thr Gln Glu Glu Glu Arg Pro

CTG GAA AGG CCA GCG AGG CTA AGA AAG GAC ACA GTT AAA CAG GCG CAG GAG
Leu Glu Arg Pro Ala Arg Leu Arg Lys Asp Thr Val Lys Gln Ala Gln Glu

GAG AGA AAG TAC TTT CTT CTC TCC GTA GAC TAT GAT GGT AAA ATG GGG AAG
Glu Arg Lys Tyr Phe Leu Leu Ser Val Asp Tyr Asp Gly Lys Met Gly Lys

GCT GTC TGC AAG CTT TAT GAT CCT GAA ACG GGT GAG CTA CAC GTC CTT TAC
Ala Val Cys Lys Leu Tyr Asp Pro Glu Thr Gly Glu Leu His Val Leu Tyr

GAC AGC ACG GGT CAC AAG TCA TAC TTC CTT GTG GAT TTA GAG CCA GAT CAG
Asp Ser Thr Gly His Lys Ser Tyr Phe Leu Val Asp Leu Glu Pro Asp Gln

ATC CAA AAA ATT CCA AAG ATT GTT AAG GAT GAG TCC TTT GTT AGG CTT GAG
Ile Gln Lys Ile Pro Lys Ile Val Lys Asp Glu Ser Phe Val Arg Leu Glu

AAG ACC ACT AAA ATA GAC CCC TAC ACT TGG AAA CCT ATT AAC CTA ACC AAG
Lys Thr Thr Lys Ile Asp Pro Tyr Thr Trp Lys Pro Ile Asn Leu Thr Lys

ATT GTG GTG AAT GAC CCC CTC GCT GTG AGA CGC CTA AGA GAA TAT GTC CCA
Ile Val Val Asn Asp Pro Leu Ala Val Arg Arg Leu Arg Glu Tyr Val Pro

AGG GCC TAT GAA GCT CAT ATA AAA TAT TTT AAC AAT TAT ATT TAC GAT TTC
Arg Ala Tyr Glu Ala His Ile Lys Tyr Phe Asn Asn Tyr Ile Tyr Asp Phe

16/28

Fig. 4 CONTINUED

AGC CTC ATA CCA GGG ATG CCC TAC GTG GTA AAG AAG GGG AAG CTA GTC CCC
Ser Leu Ile Pro Gly Met Pro Tyr Val Val Lys Lys Gly Lys Leu Val Pro

CTT AAG CCG GAG GTT GAC GTC AAA GAG CTA AAG GAA GCG TTC AAG GAT GCT
Leu Lys Pro Glu Val Asp Val Lys Glu Val Lys Glu Ala Phe Lys Asp Ala

GAC CAG ATA GCT CAA GAG ATG GCG CTA GAC TGG GCT CCC CTC TTT GAG TCC
Asp Gln Ile Ala Gln Glu Met Ala Leu Asp Trp Ala Pro Leu Phe Glu Ser

GAG ATT CCG TCG GTG AAG AGG GTC GCA ATA GAT ATA GAG GTT TAT ACT CCC
Glu Ile Pro Ser Val Lys Arg Val Ala Ile Asp Ile Glu Val Tyr Thr Pro

ATG ATG GGT AGG GTA CCG GAT CCA GTA AAG GCC GAG TAC CCC GTG ATA AGC
Met Met Gly Arg Val Pro Asp Pro Val Lys Ala Glu Tyr Pro Val Ile Ser

GTA GCC CTA GCA GGG AGC GAT GGC CTG AAA CTG GTC CTA GTC CTT GAT AGG
Val Ala Leu Ala Gly Ser Asp Gly Leu Lys Leu Val Leu Val Leu Asp Arg

GGA GAT AGT CCG ATT CAA AGT AAG GAT ATC AAG GTT GAG GTC TTC CGC ACA
Gly Asp Ser Pro Ile Gln Ser Lys Asp Ile Lys Val Glu Val Phe Arg Thr

GAG AGG GAG CTT CTC TCC AGG TTG TTT GAC ATT CTT AAG GAA TAT CCC ATG
Glu Arg Glu Leu Leu Ser Arg Leu Phe Asp Ile Leu Lys Glu Tyr Pro Met

GTT CTG ACC TTT AAC GGA GAC GAC TTC GAT ATC CCA TAC CTG ATC TTC AGA
Val Leu Thr Phe Asn Gly Asp Asp Phe Asp Ile Pro Tyr Leu Ile Phe Arg

GGT TTC AAG CTC GGG TTA CTA CAG GAT GAG ATA CCC TTC GAG ATC TCT AGT
Gly Phe Lys Leu Gly Leu Leu Gln Asp Glu Ile Pro Phe Glu Ile Ser Ser

TTT GGC AGG AAA CCT GAC GCG AAG TTC AGA TAT GGA TTT CAC ATA GAT TTG
Phe Gly Arg Lys Pro Asp Ala Lys Phe Arg Tyr Gly Phe His Ile Asp Leu

17/28

Fig. 4 Continued

TAC AGG TTC TTC TTC AAC AAG GCG GTC AGG AAC TAT GCA TTT GAG GGG AAG
Tyr Arg Phe Phe Phe Asn Lys Ala Val Arg Asn Tyr Ala Phe Glu Gly Lys

TAC TCA GAG TAC AAC CTT GAC ACC GTA GCC CAG GCA CTC TTG GGT CTC TCC
Tyr Ser Glu Tyr Asn Leu Asp Thr Val Ala Gln Ala Leu Leu Gly Leu Ser

AAG GTC AAG TTG GAC GAG TCC ATT AGC GAC CTA AAC ATG TCT AAA CTC GTG
Lys Val Lys Leu Asp Glu Ser Ile Ser Asp Leu Asn Met Ser Lys Leu Val

GAG TAC AAC TAC AGG GAC TCG GAG ATC ACG CTG AAG TTG ACC ACG TTC AAC
Glu Tyr Asn Tyr Arg Asp Ser Glu Ile Thr Leu Lys Leu Thr Thr Phe Asn

AAC GAA CTA GTA TGG AAG TTG ATT GTA CTC TTC TCC AGA ATT TCC AAG CTT
Asn Glu Leu Val Trp Lys Leu Ile Val Leu Phe Ser Arg Ile Ser Lys Leu

GGT ATA GAG GAG CTA ACT AGG ACA GAG ATA TCA GCC TGG GTA AAG AAC CTG
Gly Ile Glu Glu Leu Thr Arg Thr Glu Ile Ser Ala Trp Val Lys Asn Leu

TAC TAC TGG GAA CAT AGG AAA AGG AAC TGG TTA ATC CCC CTC AAG GAG GAA
Tyr Tyr Trp Glu His Arg Lys Arg Asn Trp Leu Ile Pro Leu Lys Glu Glu

ATC CTT GAA CGC TCC TCT GGG TTG AAG ACA GCT GCC ATT ATC AAG GGA AAG
Ile Leu Glu Arg Ser Ser Gly Leu Lys Thr Ala Ala Ile Ile Lys Gly Lys

GGA TAC AAG GGC GCA GTG GTC ATA GAC CCA CCT GTG GGG GTT TAC TTT GAC
Gly Tyr Lys Gly Ala Val Val Ile Asp Pro Pro Val Gly Val Tyr Phe Asp

GTA GTT GTT CTG GAC TTC GCC TCA CTG TAT CCC TCC ATC ATC AGG AAC TGG
Val Val Val Leu Asp Phe Ala Ser Leu Tyr Pro Ser Ile Ile Arg Asn Trp

AAC CTC AGT TAT GAA ACC GTT GAT GTG AAG GAA TGT AAC AAG AAA AGG GAT
Asn Leu Ser Tyr Glu Thr Val Asp Val Lys Glu Cys Asn Lys Lys Arg Asp

ATA AGG GAT GAG AGT GGG GCG AAA ATC CAT GAG GTG TGC GTG GAC AGG CCC

18/28

Fig. 4 CONTINUED

Ile Arg Asp Glu Ser Gly Ala Lys Ile His Glu Val Cys Val Asp Arg Pro

GGG ATT ACT GCA GTG GTA ACT GGC TTA CTT AGG GAC TTC AGG GTC AAA ATT
Gly Ile Thr Ala Val Val Thr Gly Leu Leu Arg Asp Phe Arg Val Lys Ile

TAC AAG AAG AAA GGG AAA CAG AGC AAC ATA GAC GAG GAG AGA AAG ATG TTG
Tyr Lys Lys Lys Gly Lys Gln Ser Asn Ile Asp Glu Glu Arg Lys Met Leu

TAC GAC GTG GTA CAG AGG GGC ATG AAG GTG TTC ATT AAT GCG ACC TAT GGC
Tyr Asp Val Val Gln Arg Gly Met Lys Val Phe Ile Asn Ala Thr Tyr Gly

GTC TTC GGT GCG GAG ACC TTC CCC TTG TAC GCC CCA GCA GTT GCA GAG AGC
Val Phe Gly Ala Glu Thr Phe Pro Leu Tyr Ala Pro Ala Val Ala Glu Ser

GTT ACA GCC CTA GGT AGG TAC GTA ATC ACG TCC ACC AAG GAA ATG GCT AAC
Val Thr Ala Leu Gly Arg Tyr Val Ile Thr Ser Thr Lys Glu Met Ala Asn

AAG CTT GGG CTG AAG GTT GTG TAC GGG GAT ACG GAC TCG CTC TTC ATT CAC
Lys Leu Gly Leu Lys Val Val Tyr Gly Asp Thr Asp Ser Leu Phe Ile His

CAG CCT GAT AAG AAG AAG CTG GAG GAA CTG GTG GAG TGG ACC AGG CAG AAC
Gln Pro Asp Lys Lys Lys Leu Glu Glu Leu Val Glu Trp Thr Arg Gln Asn

TTC GGG CTT GAT CTA GAG GTG GAC AAA ACT TAC AGG TTC ATT GCC TTC TCC
Phe Gly Leu Asp Leu Glu Val Asp Lys Thr Tyr Arg Phe Ile Ala Phe Ser

GGT CTT AAG AAG AAC TAC TTC GGT GTG TTC AAG GAT TCC AAG GTT GAC ATA
Gly Leu Lys Lys Asn Tyr Phe Gly Val Phe Lys Asp Ser Lys Val Asp Ile

AAG GGC ATG TTG GCA AAG AAG AGG AAC ACC CCA GAG TTT CTG AAG CAG GCC
Lys Gly Met Leu Ala Lys Lys Arg Asn Thr Pro Glu Phe Leu Lys Gln Ala

TTC AAT GAG GCT AAG GAG AGG CTA GCG AAG GTT CAG AAC CAG GAG GAG CTC
Phe Asn Glu Ala Lys Glu Arg Leu Ala Lys Val Gln Asn Gln Glu Glu Leu

19/28

Fig. 4. Continued

GAA AAG GCA ATT CAA GAC TTA ACG GCG CAG GTT AAG GAG GTG TAC AGG AAG
Glu Lys Ala Ile Gln Asp Leu Thr Ala Gln Val Lys Glu Val Tyr Arg Lys

CTT AAG ATG AAG GAA TAT AAC TTG GAT GAG CTC GCC TTC AGG GTC ATG TTA
Leu Lys Met Lys Glu Tyr Asn Leu Asp Glu Leu Ala Phe Arg Val Met Leu

TCC AGG GAC GTG AAG TCC TAT GAG AAG AAC ACC CCA CAG CAC GTT AAG GCT
Ser Arg Asp Val Lys Ser Tyr Glu Lys Asn Thr Pro Gln His Val Lys Ala

GCG GCA CAG CTG GCG GAG ATG AAC GTA CAA GTG ATG TCA AGG GAT ATA ATT
Ala Ala Gln Leu Ala Glu Met Asn Val Gln Val Met Ser Arg Asp Ile Ile

AGC TTC GTA AAG GTA AAG ACT AAG GAG GGA GTT AAA CCT GTC CAG CTA GCT
Ser Phe Val Lys Val Lys Thr Lys Glu Gly Val Lys Pro Val Gln Leu Ala

AAG CTT TCA GAG ATT GAT GTG GAT AAA TAC TAT GAG AGC GTG AGA AGT ACC
Lys Leu Ser Glu Ile Asp Val Asp Lys Tyr Tyr Glu Ser Val Arg Ser Thr

TTC GAA CAG TTA TTG AAA AGC TTC AAT GTG AGC TGG GAT AGA ATA GAG TCC
Phe Glu Gln Leu Leu Lys Ser Phe Asn Val Ser Trp Asp Arg Ile Glu Ser

2634

ACG ACA TCA ATC GAC TCG TTC TTC AAG ACT TAG
Thr Thr Ser Ile Asp Ser Phe Phe Lys Thr End

FIGURE 5

Desulfurococcus M11 TL DNA Polymerase (29PY1)

1

(SEQ ID NO:9) ATG GAG AGG GTT CGC CTA GTG AAG GTG GTT ACC AAG GAT
CCT CTA ATC GTG

(SEQ ID NO:10) Met Glu Arg Val Arg Leu Val Lys Val Val Thr Lys Asp
Pro Leu Ile Val

AGG AAG ATT AGG AGC AAG TTT AAC ACT GCG TGG GAG GCT AAG ATA AAG TAT
Arg Lys Ile Arg Ser Lys Phe Asn Thr Ala Trp Glu Ala Lys Ile Lys Tyr

CAT GCA AAC TAC ATC TAC GAT AAT AGG CTG ATA CCT GGA ATG AGG TAT GTT
His Ala Asn Tyr Ile Tyr Asp Asn Arg Leu Ile Pro Gly Met Arg Tyr Val

ACA GAC TTC TCC AAC GGT GCG CAA AAG CTT GTA ATG GTT AAG CCA GAG ATA
Thr Asp Phe Ser Asn Gly Ala Gln Lys Leu Val Met Val Lys Pro Glu Ile

CCC CAA TCC CTT GTT GAG AAA GTA AGG GAG TTG TTC AGG AAT GAG CCT CCT
Pro Gln Ser Leu Val Glu Lys Val Arg Glu Leu Phe Arg Asn Glu Pro Pro

GAA ACA GTG AAG CTG GCT GAG GAA CTC CTC CTC TTG TTC GAG GAG TCA CCG
Glu Thr Val Lys Leu Ala Glu Glu Leu Leu Leu Leu Phe Glu Glu Ser Pro

CCC AAG GTG AAG CGC GTA GCA GTC GAC ATA GAG GTT TTC ACC CCA TTC AAA
Pro Lys Val Lys Arg Val Ala Val Asp Ile Glu Val Phe Thr Pro Phe Lys

GGG CGT ATC CCC AGC CCG AAG CTC GCC GAA TAC CCT GTG ATT AGC ATA GCA
Gly Arg Ile Pro Ser Pro Lys Leu Ala Glu Tyr Pro Val Ile Ser Ile Ala

TTG GCC GGT AGC GAC GGC TTG AAG AAA ATC CTC CTG CTG GCC AGG GAA TAC
Leu Ala Gly Ser Asp Gly Leu Lys Lys Ile Leu Leu Leu Ala Arg Glu Tyr

AAG CAT GAT TTC GAC TAC ATG ATG GAG GAT TAC CCT GTT GAA GCC GAG GTG
Lys His Asp Phe Asp Tyr Met Met Glu Asp Tyr Pro Val Glu Ala Glu Val

21/28

Fig. 5 Continued - -

GAG GTG TTC GAC TCC GAG AAA GAC ATG TTG CTG GAA GCC TTC AGA ATA ATG
Glu Val Phe Asp Ser Glu Lys Asp Met Leu Leu Glu Ala Phe Arg Ile Met

GGG AGC TAT CCC GTC GTC CTC ACT TAC AAC GGT GAT AAT TTC GAC CTT CAA
Gly Ser Tyr Pro Val Val Leu Thr Tyr Asn Gly Asp Asn Phe Asp Leu Gln

TAC CTG TAC GTG AGA GCC TTC AAG CTG GGG ATT CTG AGA AGC CAT ATC CCG
Tyr Leu Tyr Val Arg Ala Phe Lys Leu Gly Ile Leu Arg Ser His Ile Pro

TTG AAG ATA GGG GAG GAT ATG ATT AGA ATT GAC ACA AGC ATA CAC CTA GAT
Leu Lys Ile Gly Glu Asp Met Ile Arg Ile Asp Thr Ser Ile His Leu Asp

CTA TAC AAG TTC TTC TCG AAC AGG GCT GTT AAA AAC TAT GCT TTC GGG GGG
Leu Tyr Lys Phe Phe Ser Asn Arg Ala Val Lys Asn Tyr Ala Phe Gly Gly

AAA TAC CAG GAG GAG AAG CTT GAC GCT GTT TCA GGG GCA CTG CTA GGA GTG
Lys Tyr Gln Glu Glu Lys Leu Asp Ala Val Ser Gly Ala Leu Leu Gly Val

TCG AAA ATA GGT TTC GAG GAA ACA ATC GGC GGC ATA TCG GCC TCA CTA TTA
Ser Lys Ile Gly Phe Glu Glu Thr Ile Gly Gly Ile Ser Ala Ser Leu Leu

GCC GCC TAC AAC TAC AGG GAT GCC GAG ATC ACG TTA AAC CTA ACC ATG TTC
Ala Ala Tyr Asn Tyr Arg Asp Ala Glu Ile Thr Leu Asn Leu Thr Met Phe

AGT AAT GAA CTC GTT TGG AAA CTC ATT ATT CTT CTA GCT AGG GTT TCC AAG
Ser Asn Glu Leu Val Trp Lys Leu Ile Ile Leu Leu Ala Arg Val Ser Lys

ACA AGC ATT GAA GAC CTG TGT AGG AGG CAG ATT TCC TAC TGG ATT CAA AAT
Thr Ser Ile Glu Asp Leu Cys Arg Arg Gln Ile Ser Tyr Trp Ile Gln Asn

CTG TTC TTC TGG GAG CGC AGG AAG CTC GGC TAC CTC ATA CCT AAC AAG GAG
Leu Phe Phe Trp Glu Arg Arg Lys Leu Gly Tyr Leu Ile Pro Asn Lys Glu

22/28

Fig. 5 CONTINUED - -

GAC ATT CTG AGG CAT GTA AGG GGG ACG GGG ACG AAG GCG ATT ATT GAG GGT
Asp Ile Leu Arg His Val Arg Gly Thr Gly Thr Lys Ala Ile Ile Glu Gly

AAG AAG TAC GGT GGA GCC TTA GTG GTT GAG CCT CCG AAA GGG GCT TTC TTC
Lys Lys Tyr Ala Gly Ala Leu Val Val Glu Pro Pro Lys Gly Ala Phe Phe

AAC GTG GTC GTC CTC GAC ATA GCC TCC CTA TAC CCT AGC ATT ATC AAA AAA
Asn Val Val Val Leu Asp Ile Ala Ser Leu Tyr Pro Ser Ile Ile Lys Lys

TAC AAT CTG AGC TAT GAG ACC GTT GAC ATG AAG TGG TGT AGC AAG ACA ATA
Tyr Asn Leu Ser Tyr Glu Thr Val Asp Met Lys Trp Cys Ser Lys Thr Ile

GAT ATT GTC GAT GAA ACC GGG AGA AGG CTT CAC GAA GTC TGC GTT GAC AAG
Asp Ile Val Asp Glu Thr Gly Arg Arg Leu His Glu Val Cys Val Asp Lys

CCC GGG TTG ACC GCG CAA CTA ACC GGT ATT CTA AGG GAT TAC AGG GTT GGA
Pro Gly Leu Thr Ala Gln Leu Thr Gly Ile Leu Arg Asp Tyr Arg Val Gly

ATA TAT AAG AAG AGG TCT AAG GAT AAG AGC CTT CCC CCT GAA ACC CTG GCC
Ile Tyr Lys Lys Arg Ser Lys Asp Lys Ser Leu Pro Pro Glu Thr Leu Ala

TGG TAC GAG GTG GTT CAG AGA GCT ATT AAG GTG TTC ATA AAC GCT AGC TAC
Trp Tyr Glu Val Val Gln Arg Ala Ile Lys Val Phe Ile Asn Ala Ser Tyr

GGG GTC TTC GGG GAT GAG AAG TTC TCT CTG TAC TCC CCA GCA GTG GCT GAA
Gly Val Phe Gly Asp Glu Lys Phe Ser Leu Tyr Ser Pro Ala Val Ala Glu

AGC GTT ACC GCG ATG GGT AGG AAG TCC TTC TAC ACT ATT GTG AGA AAG GCC
Ser Val Thr Ala Met Gly Arg Lys Ser Phe Tyr Thr Ile Val Arg Lys Ala

GCG GAT CTC GGG GTT AAA ACA CTG TAT GGC GAC ACG GAC TCG ATA TTC GTC
Ala Asp Leu Gly Val Lys Thr Leu Tyr Gly Asp Thr Asp Ser Ile Phe Val

TGG GCC CCA ACC CAG GAG CAG TTG AGG AAG CTA CAG TCA TGG ATC CTT GAG

23/28

Fig 5 CONTINUED

Trp Ala Pro Thr Gln Glu Gln Leu Arg Lys Leu Gln Ser Trp Ile Leu Glu

AAG CTA GGC CTG GAG ATC GAG ATT GAC AAG TCT TTT ACA TAC GTG GTT TTC
Lys Leu Gly Leu Glu Ile Glu Ile Asp Lys Ser Phe Thr Tyr Val Val Phe

ACA GGG CTT AAG AAG AAC TAC CTG GGC AGA ACG GTT GAC GGC GGC ATA GAG
Thr Gly Leu Lys Lys Asn Tyr Leu Gly Arg Thr Val Asp Gly Gly Ile Glu

ATC AAG GGG CTT GTC SCG AAG AAG AGG AAT ACT CCG GAG TTC CTG AAA GAC
Ile Lys Gly Leu Val Xxx Lys Lys Arg Asn Thr Pro Glu Phe Leu Lys Asp

TTG TTC GAG AAT GTT ATC GAA AAG CTT AAA AGC GTT GAA AAC CCC GCG GGT
Leu Phe Glu Asn Val Ile Glu Lys Leu Lys Ser Val Glu Asn Pro Ala Gly

TTC ATA GAG TTC GTC AAG TGG TTG GAG CAT CAG GTG AAG ACA ATA CAT AAC
Phe Ile Glu Phe Val Lys Trp Leu Glu His Gln Val Lys Thr Ile His Asn

GAT ATT AGG AGG AAG GAG ATA ACG CTC GAC CGG CTC GCC ATA AGG GTG GCC
Asp Ile Arg Arg Lys Glu Ile Thr Leu Asp Arg Leu Ala Ile Arg Val Ala

TTA ACC AAG ACG CCA TCC CTC TAC ACT AAG ACT AAG CCG CCG CAT GTT AAG
Leu Thr Lys Thr Pro Ser Leu Tyr Thr Lys Thr Lys Pro Pro His Val Lys

GCA GCC CTC CAA TTA ATG AAC TAC GGG TAC AGC GTG GAG GAG GGG GAT ATT
Ala Ala Leu Gln Leu Met Asn Tyr Gly Tyr Ser Val Glu Glu Gly Asp Ile

ATA ACG TTT GTC AAG GTG AAG AGC AAG GAG GGC TAT AAG GCT ATA CAG TTA
Ile Thr Phe Val Lys Val Lys Ser Lys Glu Gly Tyr Lys Ala Ile Gln Leu

ACG AGG CTT CAC GAA GTA GAC CCT GAT AAG TAC ATT GAG CTT GTT AAA AGC
Thr Arg Leu His Glu Val Asp Pro Asp Lys Tyr Ile Glu Leu Val Lys Ser

GGT CTT GAA CAA TTC CTC TCA GCC TTC GGA ATA AGG TGG GAG GAT ATC ATA
Gly Leu Glu Gln Phe Leu Ser Ala Phe Gly Ile Arg Trp Glu Asp Ile Ile

24/28

Fig. 5 Continued --

2289

GGC TCC GGC GGG TTA ACC GAG CTT TTG AGA AAC AAT AGG GCG TAG

Gly Ser Gly Gly Leu Thr Glu Leu Leu Arg Asn Asn Arg Ala End

25/28

FIGURE 6

Aquifex VF-5 DNA Polymerase (34PY)

1

(SEQ ID NO:11) ATG GAT TTT GAA TAC GTA ACG GGA GAA GAG GGA TTA AAA AAG
GCA A¹ AAA

(SEQ ID NO:12) Met Asp Phe Glu Tyr Val Thr Gly Glu Glu Gly Leu Lys
Lys Ala Ile Lys

AGG CTC GAA AAT TCT CCA TAC CTT TAC CTG GAT ACG GAA ACC ACA GGA GAC
Arg Leu Glu Asn Ser Pro Tyr Leu Tyr Leu Asp Thr Glu Thr Thr Gly Asp

AGG ATA AGG CTC GTA CAA ATC GGA GAC GAA GAA AAC ACC TAC GTT ATT GAC
Arg Ile Arg Leu Val Gln Ile Gly Asp Glu Glu Asn Thr Tyr Val Ile Asp

CTC TAC GAA ATT CAG GAT ATA GAA CCT CTG AGG AAA TTA ATA AAC GAA AGG
Leu Tyr Glu Ile Gln Asp Ile Glu Pro Leu Arg Lys Leu Ile Asn Glu Arg

GGG ATA GTA GGG CAC AAC CTT AAG TTC GAT CTT AAG TAC CTC TAC AGG TAC
Gly Ile Val Gly His Asn Leu Lys Phe Asp Leu Lys Tyr Leu Tyr Arg Tyr

GGG ATA TTT CCC TCG GCA ACG TTT GAC ACT ATG ATA GCG AGC TAC CTC CTC
Gly Ile Phe Pro Ser Ala Thr Phe Asp Thr Met Ile Ala Ser Tyr Leu Leu

GGA TAC GAG AGA CAC TCC CTC AAT CAC ATA GTT TCA AAC CTA CTC GGA TAT
Gly Tyr Glu Arg His Ser Leu Asn His Ile Val Ser Asn Leu Leu Gly Tyr

TCC ATG GAC AAG AGT TAT CAG ACT TCC GAC TGG GGA GCG AGC GTT CTG AGC
Ser Met Asp Lys Ser Tyr Gln Thr Ser Asp Trp Gly Ala Ser Val Leu Ser

GAC GCT CAG CTC AAG TAC GCT GCA AAC GAC GTT ATA GTC CTC AGA GAA CTC
Asp Ala Gln Leu Lys Tyr Ala Ala Asn Asp Val Ile Val Leu Arg Glu Leu

TTC CCT AAG ATG AGG GAC ATG TTA AAC GAG CTA GAC GCT GAG AGG GGA GAG
Phe Pro Lys Met Arg Asp Met Leu Asn Glu Leu Asp Ala Glu Arg Gly Glu

Fig 6 Continued - -

GAA CTG CTC AAG ACT AGA ACG GCA AAG ATT TTC GAT CTG AAG AGT CCC GTA
Glu Leu Leu Lys Thr Arg Thr Ala Lys Ile Phe Asp Leu Lys Ser Pro Val

GCA ATA GTG GAA ATG GCT TTC GTA AGG GAA GTT GCA AAA CTC GAG ATA AAC
Ala Ile Val Glu Met Ala Phe Val Arg Glu Val Ala Lys Leu Glu Ile Asn

GGC TTT CCC GTG GAC GTA GAA GAG CTA ACC AAC AAG TTA AAA GCT GTG GAA
Gly Phe Pro Val Asp Val Glu Glu Leu Thr Asn Lys Leu Lys Ala Val Glu

AGG GAA ACC CAG AAG AGG ATA CAG GAG TTT TAC ATA AAG TAC AGA GTT GAC
Arg Glu Thr Gln Lys Arg Ile Gln Glu Phe Tyr Ile Lys Tyr Arg Val Asp

CCT CTC TCT CCG AAA CAG CTC GCC TCA CTC CTG ACG AAG AAG TTT AAA CTG
Pro Leu Ser Pro Lys Gln Leu Ala Ser Leu Leu Thr Lys Lys Phe Lys Leu

AAC CTT CCC AAG ACT CCT AAA GGG AAC GTA TCT ACA GAC GAC AAG GCT CTT
Asn Leu Pro Lys Thr Pro Lys Gly Asn Val Ser Thr Asp Asp Lys Ala Leu

ACT TCC TAT CAG GAC GTA GAA CCC GTA AAA CTC GTT CTG GAA ATA AGA AAG
Thr Ser Tyr Gln Asp Val Glu Pro Val Lys Leu Val Leu Glu Ile Arg Lys

CTT AAG AAG ATC GCG GAC AAG TTA AAG GAG TTA AAA GAA CAC TTG AAG AAC
Leu Lys Lys Ile Ala Asp Lys Leu Lys Glu Leu Lys Glu His Leu Lys Asn

GGG AGA GTT TAC CCG GAG TTC AAG CAA ATA GGA GCT GTA ACG GGA AGG ATG
Gly Arg Val Tyr Pro Glu Phe Lys Gln Ile Gly Ala Val Thr Gly Arg Met

TCC TCC GCA CAC CCA AAT ATC CAG AAC ATA CAC AGG GAT ATG AGA GGA ATT
Ser Ser Ala His Pro Asn Ile Gln Asn Ile His Arg Asp Met Arg Gly Ile

TTC AAG GCG GAG GAG GGA AAT ACT TTC GTC ATT TCG GAC TTT TCT CAG ATA
Phe Lys Ala Glu Glu Gly Asn Thr Phe Val Ile Ser Asp Phe Ser Gln Ile

27/28

Fig 6 CONTINUED - -

GAG CTC AGG ATT GCG GCC GAA TAC GTA AAG GAC CCG CTT ATG CTG GAC GCC
Glu Leu Arg Ile Ala Ala Glu Tyr Val Lys Asp Pro Leu Met Leu Asp Ala

TTC AAA AAG GGA AAG GAC ATG CAC AGG TAC ACC GCT TCA GTG GTA CTC GGA
Phe Lys Lys Gly Lys Asp Met His Arg Tyr Thr Ala Ser Val Val Leu Gly

AAG AAA GAG GAA GAA ATA ACA AAA GAG GAG AGA CAG CTC GCA AAA GCT ATA
Lys Lys Glu Glu Glu Ile Thr Lys Glu Glu Arg Gln Leu Ala Lys Ala Ile

AAC TTC GGT CTC ATA TAC GGC ATT TCC GCT AAA GGG CTT GCA GAA TAC GCA
Asn Phe Gly Leu Ile Tyr Gly Ile Ser Ala Lys Gly Leu Ala Glu Tyr Ala

AAG CTT GGT TAC GGC GTT GAA ATT TCT TTA GAA GAA GCT CAG GTT TTG AGA
Lys Leu Gly Tyr Gly Val Glu Ile Ser Leu Glu Glu Ala Gln Val Leu Arg

GAG AGG TTT TTC AAG AAC TTC AAA GCT TTC AAA GAG TGG CAC GAC AGA GTT
Glu Arg Phe Phe Lys Asn Phe Lys Ala Phe Lys Glu Trp His Asp Arg Val

AAG AAA GAA CTA AAG GAA AAG GGA GAG GTA AAA GGT CAT ACG CTT CTT GGA
Lys Lys Glu Leu Lys Glu Lys Gly Glu Val Lys Gly His Thr Leu Leu Gly

AGG AGA TTT TCC GCA AAT ACC TTT AAC GAC GCT GTA AAT TAC CCC ATA CAG
Arg Arg Phe Ser Ala Asn Thr Phe Asn Asp Ala Val Asn Tyr Pro Ile Gln

GGA ACG GGT GCG GAC CTA CTA AAA CTG GCA GTT CTA CTT TTT GAC GCA AAC
Gly Thr Gly Ala Asp Leu Leu Lys Leu Ala Val Leu Leu Phe Asp Ala Asn

CTC CAG AAA AAG GGA ATA GAT GCA AAG CTC GTG AAC CTC GTG CAC GAC GAG
Leu Gln Lys Lys Gly Ile Asp Ala Lys Leu Val Asn Leu Val His Asp Glu

ATA GTC GTA GAG TGC GAA AAG GAA AAA GCG GAA GAA GTA AAA GAA ATA CTC
Ile Val Val Glu Cys Glu Lys Glu Lys Ala Glu Glu Val Lys Glu Ile Leu

GAA AAA TCC ATG AAA ACG GCG GGA AAG ATA ATA CTG AAA GAG GTT CCC GTG

28/28

Fig. 6 Continued--

Glu Lys Ser Met Lys Thr Ala Gly Lys Ile Ile Leu Lys Glu Val Pro Val

1725

GAA GTA GAA AGC GTT ATA AAC GAA AGG TGG ACG AAA GAT TAA

Glu Val Glu Ser Val Ile Asn Glu Arg Trp Thr Lys Asp End

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US98/17152

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : C12N 9/12, 15/54; C07K 16/00; C07H 21/04

US CL : 435/194, 252.3, 320.1; 536/23.2, 24.32; 530/387.1

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/194, 252.3, 320.1; 536/23.2, 24.32; 530/387.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

NONE

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US H1531 A (BLUMENTALS et al.) 07 May, 1996, see entire document.	1-30



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
A document defining the general state of the art which is not considered to be of particular relevance	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
B earlier document published on or after the international filing date	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*Z* document member of the same patent family
O document referring to an oral disclosure, use, exhibition or other means	
P document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

02 OCTOBER 1998

Date of mailing of the international search report

29 OCT 1998

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

KEITH D. HENDRICKS

Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

national application No.
PCT/US98/17152

B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

APS, DIALOG

search terms: DNA polymerase, Ammonifex degensii, Pyrolobus fumarius, Archaeoglobus lithotrophicus, Metallosphaera
prunae, Desulfurococcus, Aquifex VF-5.